Conference Papers

4th Sub-Regional Conference on Assessment in Education

Hosted by Umalusi

At the of University of Johannesburg, School of Tourism and Hospitality

26th to 30th June 2006

PUBLISHED BY

UMALUSI COUNCIL FOR QUALITY ASSURANCE IN GENERAL AND FURTHER EDUCATION AND TRAINING

37 General Van Ryneveld Street, Persequor Technopark, Pretoria
Telephone: 27 12 3491510 • Fax: 27 12 3491511 • info@umalusi.org.za

Acknowledgements

Umalusi would like to thank our printers, Idube Security Printers, for generously covering the cost of printing this book, as well as conference t-shirts, invitations to conference events, and conference certificates of attendance.

Our other generous sponsors are listed on the insert, and we would like to express our extreme gratitude to all of them.

Umalusi would also like to express its gratitude to Shireen Badat and Matseleng Allais, for their energy, vision, determination, sense of humour, attention to detail, creativity, and ceaseless hard work, in preparing for this conference. They were assisted by Ramadimetja Phaladi, Frank Chinyamakobvu, and Eugene du Plooy in the build up to the conference, and we would also like to thank them for their hard work and dedication.

Table of contents

Introduction

Umalusi is very proud to host the 4th Annual Sub-Regional Conference on Assessment in Education. Since the first conference took place four years ago in Zambia, the conference has proved to be an extremely useful forum for Examinations Councils and other organizations involved in educational assessment in Southern African countries to share ideas, best practice, challenges, and solutions. We hope this conference will continue this emerging tradition, and consolidate a stronger working relationship between the countries present.

The theme of the 2006 conference is *Assessment and Educational Standards*. Umalusi felt that this was an appropriate theme given the international increasing concern about 'standards'. There is ongoing public debate about educational standards being too high or too low, and assessment and quality assurance bodies are confronted with criticism from the public if pass marks go up or down. But what is meant by standards? How are they determined? Who plays a role in determining them? What is the role of assessment? How does it relate to the new standards-based educational reform movement, and outcomes-based education, if at all? How should the standard of African education systems relate to international standards?

'Standards' is a crucial, but difficult issue for assessment bodies. We felt that the time is right to share ideas about how some of these issues have been tackled in our different systems. This conference is an opportunity for members of examinations boards in SADC countries to share research, systems, approaches, and ideas about standards, systems of standards setting and monitoring, and systems to ensure that educational programmes are relevant and appropriate. The conference has various sub-themes, which explore the notion of standards in different ways:

- Thinking about standards over time, across subjects, and across learning pathways
- Standards-based and outcomes-based educational reform
- Determining standards
- Measuring learner performance against standards
- International standards and local knowledge.

As can be seen from this conference package, we have received a range of excellent papers addressing these themes, from nearly all countries in the Southern African sub-region. This book contains all except three papers that will be presented at the conference, and we hope that it is useful to delegates, as they listen to papers being presented, but that it also enables them to share the conference papers with others in their organizations, who did not manage to attend the conference.

Programme

JUNE 25TH (SUNDAY)

Reception for delegates at 6pm, at Milpark Garden Court. Registration for SADC delegates and out-of-town South African delegates from 6-7 at the Milpark Garden Court Hotel.

DAY ONE: JUNE 26TH (MONDAY)

Time	Session	Speakers
07:30	**Registration and tea**	
09:00	*Opening session*	Opening: Dr Peliwe Lolwana, Umalusi
	Chair: South Africa	Greeting: Dr Ihron Rensburg, University of Johannesburg
		Greeting and handover of chairpersonship to South Africa: MW Matemba, Malawi National Examinations Board
		Invocation
		Keynote address: Prof Crain Soudien, University of Cape Town: *Making our own modernity: standards, values, and benchmarks in the African school in the age of globalization*
11:30	Tea	
12:00	*Thinking about standards over time, across subjects, and across learning pathways*	Dr Nick Taylor, Joint Education Trust: *Accountability and support in school development in South Africa*
	Chair: Swaziland	Mr David Yadidi, Malawi National Examinations Board: *MSCE Examination results—Are they a reliable tool for measuring educational standards?*
13:00	Group photograph	
13:30	Lunch	
14:30	*Thinking about standards over time, across subjects, and across learning pathways*	Ms Ronel Blom, SAQA: *Parity of esteem between vocational and general education*
	Chair: Tanzania	Mr Gerald Chiunda, Malawi National Examinations Board: *Can we ensure that vocational programmes are at the same 'standard' as general education programmes?The case of Geometrical and Orthographic Drawing in Malawi Craft*

Time	Session	Speakers
		Ms Matseleng Allais, Umalusi: *Apples and oranges: comparing vocational and general subjects*
16:00	Tea	
16:30	*Thinking about standards over time, across subjects, and across learning pathways*	Ms Betty Mutambanengwe, Zimbabwe School Examinations Council: *Determining standards in practical subjects*
	Chair: Mozambique	Ms Surette van Staden and Prof Sarah Howie, University of Pretoria: *Assessing the proficiency of adult learners in reading literacy in South Africa*
18:00	Cocktail party and finger supper with the Minister of Education at the University of Johannesburg, School of Hospitality and Tourism	

DAY TWO: JUNE 27TH (TUESDAY)

Time	Session	Speakers
08:30	*Standards-based and outcomes-based educational reform*	Mr Joe Samuels, SAQA: *SADC Joint model for assessment and certification*
	Chair: Botswana	Prof Linda Chisholm, Human Sciences Research Council: *Diffusion of NQF and OBE in SADC*
		Mr Francis Chirume, Zimbabwe School Examinations Council: *Standards-based educational reform and implications for school-based assessment*
		Ms Cindy Ramhurry, University of Johannesburg: *A recognition of assessment practices in South Africa from a Foucauldian perspective*
10:30	Tea	
11:00	*International standards and local knowledge*	Dr Kai Horsthemke, Wits University: '*Indigenous knowledge', assessment and international standards*
	Chair: Zambia	Mr Toliwe Chehore, Zimbabwe School Examinations Council: *Ethno-based learning and international standards*
		Ms Vanessa Scherman, Ms Elizabeth Archer, and Prof Sarah Howie, University of Pretoria: *Locating the local village within the global village: assessment possibilities and practical challenges*
		Dr NgoatoTakalo, North West University: *International standards versus local curriculum objectives*
13:00	Lunch	
14:00	*Assessing learners against*	Mr Rufus Poliah, South African National Department of Education: *Constructing an integrated model for the quality*

Time	Session	Speakers
	standards	*assurance of school-based assessment in preparation for the National Senior Certificate*
	Chair: Zimbabwe	
		Mr Newton Mutanekelwa and Mr Gabriel Mweemba, Examinations Council of Zambia: *Towards piloting school-based continuous assessment at middle basic level: The Zambian approach*
		Dr Thabiso Nyabanyaba, Wits Education Policy Unit: *The fallacy of measuring mathematics 'competencies' in a context of 'high stakes' examinations*
		Mr Prince Masilo, Umalusi: *Determining standards—moderation of marking with specific reference to external moderation performed by Umalusi*
16:00	Tea	
16:30	*Determining standards/ Assessing learners against standards*	Dr Morgan Naidoo, KZN Department of Education: *Adjustment of marks and the uses of norms and standards in the Senior Certificate examination*
	Chair: Lesotho	Mr Clive Long, IEB: *Using the internet to monitor the moderation of School Based Assessment (SBA) as a first step to innovation of SBA*
18:00	Dinner and evening activity	Dinner at Moyo, Market Theatre. Buses depart from University of Johannesburg, School of Hospitality and Tourism.

DAY THREE: JUNE 28TH (WEDNESDAY)

Time	Session
08:30	Outing: social history tour of Johannesburg. Buses depart from University of Johannesburg, School of Hospitality and Tourism at 8:30.
18:00	Short tour of Constitutional Hill. Cocktail party with finger supper at the former Women's Prison, Constitutional Hill.

DAY FOUR: JUNE 29TH (THURSDAY)

Time	Session	Speaker
8:30	*Determining standards/Assessing learners against standards*	Mr Joe Cesare, Gauteng Department of Education and Prof Coert Loock, University of Johannesburg: *Remarking of examination answer scripts—finding a standard for quality assurance*
	Chair: Malawi	
		Dr Serara Moahi and Ms Magolegwa Kwele, Examinations Research and Testing Division, Ministry of Education, Botswana: *Educational Standards, What they are and why they matter. Planting the seed of a standards-led educational reform movement in Botswana*
		Ms Anne Oberholzer, IEB: *Looking for a needle in a*

Time	Session	Speaker
		haystack: Establishing 'standards' at the end of GET (grade 9)
		Mr Sandile Ndaba, Umalusi: '*A tail of many dogs': Assessment and the scramble for standards*
10:30	Tea	
11:00	*Determining standards/Assessing learners against standards*	Mr Mark Potterton, Umalusi: *Standards—a couple of steps back*
	Chair: Namibia	Dr Duncan Nyirenda, Mzuzu University: *Determining standards in Malawi*
		Ms Rosalynd Janisch, IEB: *Getting it right; keeping it right: Portfolios for Home Languages*
12:30	Lunch	
14:00	Outing: Apartheid museum. Buses depart from the University of Johannesburg, School of Hospitality and Tourism.	
18:00	Business meeting	
19:00	Gala dinner at the University of Johannesburg, School of Hospitality and Tourism.	

DAY FIVE: JUNE 30TH (FRIDAY)

Time	Session	Speaker
9:30	*Assessing learners against standards*	Dr Edith Dempster, University of KwaZulu-Natal: *Strategies for answering multiple choice questions among South African learners: What can we learn from TIMSS 2003?*
	Chair: South Africa	
		Ms Anitha Ramsuran, University of KwaZulu-Natal: *How are teachers' practices positioned in discourses of assessment?*
		Mr Chris Nyangintsimbi, Umalusi: *Lessons learnt from the evaluation of examination systems for the Senior Certificate*
11:00	Tea	
11:30	Closing session	Dr Peliwe Lolwana, Umalusi: *Does quality assurance improve the quality of education? Lessons from South Africa*
	Chair: South Africa	
		Choir
		Closing address: Deputy Director General of Education, Ms Penny Vinjevold
14:00	Lunch and departure	

Keynote address:

Making our own modernity: standards, values, and benchmarks in the African school in the age of globalization

Crain Soudien, University of Cape Town

Introduction

There are many risks in talking about Africa, not least, as recent critiques at the Klein Karoo Nasionale Fees in Oudtshoorn showed in April, the persistent tendency to treat it as a single and homogenised space (see *Cape Times,* Friday May 26, 2006:11). Africa*,* in this vein, is notoriously susceptible to hyperbole and caricature. On the one side, one has racist ideologues insistently projecting it in the alarmist language of irredeemable disease and/or savagery, while, on the other, there are starry-eyed dreamers who see it only in the innocence of the original birthplace, the alpha and omega of human existence. Neither of these is, of course, all there is to the continent. It is complex, multifaceted and distinctly not able to able to be described in single terms.

And yet, no-one will deny, Africa faces major challenges. These challenges manifest themselves across a range of areas of social life. Many of them are self-inflicted, but the blight of colonialism, and only the most blind will deny this, is deep and now wholly institutionalised as a structural and even psychological reality. In education, the burden is as great, if not more so, as may be the case in many other areas of social welfare. The general picture in African educational systems is bleak. In terms of this picture:

1. Schools are not functioning
2. Children are not performing adequately
3. Teachers appear to be failing.

It is this that we are having to confront. Using this backdrop, the purpose of this talk is to consider how we might begin to think of education in Africa, not on its own terms, because isolation from the rest of the world is virtually impossible, as the argument I will make later will hopefully show, but critically seizing hold of its own destiny.

A place to start

Interestingly, and I acknowledge that the conditions are not identical, we find ourselves in a similar position to where India was at independence in 1947. For India then, and for us now, the question was how we might build the nation. The echoes across these last sixty years are strong. In the lead up to its independence a debate was taking place between Nehru and Gandhi about modernity, modernisation and the West. At the heart of the discussion, driven by an "urge to establish a modernity of [their] own, one that differed from Western

modernity", and which was, therefore, also a critique of modernity, was the question of what the character of the new Indian nation-state should be (Gyan Prakash, 1999:201). It was here that Nehru bent one way and Gandhi another. Different as their postures were, both found themselves asking hard questions of the relationship of India's past to its present: "What was the India that was brought into view by the people in the villages and by their cries of 'Victory to the Motherland'?"

Nehru's answer to what the modern India should be was 'science' (Prakash, 1999:207). He had no wish to live in a timeless Indian past, but he thought he saw in the people "an urge driving the people in a direction not wholly realized" (Prakash, 1999:204). To fulfil this realization, what the Indian people had to do, he felt, was modernize Indian society—the disavowal of religiosity and the embrace of science. Such a science, however, had to grow out of India's national roots, which were different to that of the West which had a long way to go in developing science as "a method, an approach, a critical temper in the search for truth" (ibid:209). At the same time, in his vision of building an India that was both national and international, he was quick to rule out imitation of the West (ibid:208). The West had come to science without being able to say anything about the purpose of life. In India lay the possibility for another road to modernity—its past was not dead but alive, and ready to give direction to science and so modernity.

For Gandhi it was quite a different matter. For him it was fundamentally about discarding the Western idea of modernity. This modernity represented conflict and competition which were alien to India's tradition of village life which stood for warmth and the intimacy of family. This intimacy was built on-face-to-face relationships which bound people together in an ethical order in which mutual dependence forced self-discipline. As Prakash says, "Indians were to attain their freedom and become national subjects by discarding Western disciplines and returning to indigenous sources of self-subjectification" (ibid:219).

The eloquence of this debate is powerful and evocative as we ponder our own modernity and ask hard questions about our own choices as nations in the making or as nations attempting to locate themselves in the global order. Its echoes ring loud in the debate that is taking place in the papers that are being presented in this conference. Many of them, in a range of provocative ways, are rightly posing the question of what our standards ought to be in relation to the world. I say rightly so because if we are to become more than imitators of the divisive and hierarchalising values and approaches of the West, we are obliged to make this debate one that reaches right down to the streets and homes of our countries. The consequences of failing to do so are dire. Most of us, out of sheer convenience, will simply float along with whatever the unconscious settlement is that our societies will make. The problem with unconsciousness settlements is that they facilitate practices that become socially sedimented and once these seize hold, fascinating as objects of sociological analysis as they might be, they are extremely difficult to undo. They critically, also, signal a giving away of our right to intervene in our own histories.

In confronting our choices what the Indian precedent provides us with is an opportunity to structure our own debate. Important about its broad parameters is that it recognises the complex relationship it holds to the idea of the West. Intrinsic to this complexity is the

recognition of the Western idea of the centralised state and the acknowledgement that state-building, state-craft and the idea of the nation around the state are unavoidable responsibilities. What kind of state, however, India should build signalled the question marks posed of the state form that had emerged in the West. The discussion was, therefore, inherently a critical one.

What the options before us emphasize is how much we need to understand the full implications of the decisions we make. The rest of my discussion attempts to show what these are and the implications within each.

I suggest in the discussion that follows that we essentially have three visions with which to work in deciding what orientation we wish to take, in deciding what our own modernity will be.

The recovery of past glories: the utopians

There is an intense desire amongst many oppressed people in the world to 'escape' from the embrace of the globalised mainstream. We see this in particular versions of Afrocentric, indigenous-knowledge and other religiously-centred philosophies. An example of the first is what Adams (1997) describes as the self-determination school of thought in the United States. This view emerged out of frustration with the mainstream establishment and sought to separate black people physically and socially from the majority society and to create an independent "environment such as a state in which blacks can implement their survival strategies" (Adams, 1997:441). This frustration is also seen in the work of indigenous-knowledge proponents. An example is that of Brady (1997:421), arguing for restoring Aboriginal knowledge in Australia. She says that her ancestors had in place systems of education and social cohesion "which sustained them for 40,000 years… I believe that it is time we empowered ourselves to take back our education so that we can move with pride into the next 40,000 years." Another example is found on the African continent itself and is promoted by Banteyerga (1994) who argues that "'modern education' is not satisfactorily addressing the problems of Africa to meet the needs and aspirations of the African people." Supporting Banteyerga, Nekhwevha (1999) makes the point that Africans need to move away from "their long academic sojourn" in the Western imagination and should struggle "to make African culture and experience the primary constituent of our world view." For Nekhwevha (2000) this approach would be integrative, empowering and liberatory.

Central to all these critiques is a very specific description of the forms of globalisisng hegemony against which they are fighting. These hegemonies are predatory and have no respect for local culture and local knowledges. In globalising the local, their instinct is to instantly displace or relegate non-Western forms of understanding and knowledge to the margin. Evident in the analysis of people in this group - the utopians - is a serious and sustained critique of the hierarchalising and ranking, the dividing, and indeed the 'othering' proclivities of the West and its economic, cultural and social forms. The problems pointed to are real and deep.

Unfortunately, however, the critiques fail to engage the complexity of dominance. On the one hand, when the critique is expressed in cultural terms, whether it comes from black

feminists or religious and political fundamentalists, it fails to provide the basis for a real alternative knowledge system. While the literary outsiders and some of the African-centred philosophies, such as Audre Lorde's, assert a combativeness, the effect of their work is to produce a set of dispositions and behaviours which, in the end, turns them into critical 'insiders'. When the critique takes the form of socio-economic analysis, such as that of political-economist Samir Amin, who seeks to break from the West, it fails to engage with the politics of the conjuncture and to deal with the already existing conditions of the connectedness of Africa to the rest of the world. Absent in their approach is a recognition of the already-connected world and the ways in which institutions and individuals are, everywhere, articulated into processes and protocols which cannot, it is suggested, be conducted on the basis of a national or regional independence.

What these challenges lose sight of, therefore, without wishing to diminish the importance of their critique of mainstream education, is how Africanisation or indigenous knowledge systems are *already* engaged in articulation with the global world. It is true, as Seepe (1998) and others imply, that intense processes of cultural alienation have taken place within African communities. What, however, an appeal for reviving a *displaced* Africa underestimates is the extent to which African people continue to hold on to their own cultural practices, are taking these practices into modernity—the dominance of institutions and practices defined by ideas of rationality such as humanism, individualism, democracy, parliament, systems of justice, education and so on that emerged in Europe in the 18th century—and, are, in the process, redefining modernity, and, indeed, their own traditions. Globalisation in this situation is not a one-way process. Let us hold on to that last thought as we move to looking at the second vision embodied in new modernism.

Nehru, Mbeki and the knowledge vision: the new modernists

Part of the difficulty of the utopian ideal is that we have already conceded that we cannot live outside of the global order, and realise, much more so I think than Nehru did more than sixty years ago, that this order does indeed have the ability, as Mbeki recognised during his African Renaissance address in 2001, to influence where we go. Our very point of departure in thinking about our place in the world, our standards and our modernity, therefore, is a point of weakness. We have lost the power to declare our independence from the global order. It is impossible, as a consequence, to even contemplate Gandhi's utopia as a response to modernity. We are, for better or worse, deep inside the global system.

In this regard, the contributions made by African leaders are important. Amongst these, President Mbeki's contributions are critical because he is, arguably, paying the most attention to these questions, having repeatedly expressed himself with respect to the question of Africa's relationship to the West. What he has to say is important to engage with critically. Addressing the Third African Renaissance Festival in Durban in 2001, he opened his remarks with the following comment: "As Africans we are faced with the urgent challenge of ending poverty and underdevelopment on our Continent…. The first objective we confront… is that we must ourselves take on the responsibility to answer the question – what are the ways and means that we must adopt to ensure that we achieve these objectives" (Mbeki: 2001:1). In framing his answer he begins with the comment that

> …it is necessary that the peoples of Africa gain the conviction that they are not, and must not be wards of benevolent guardians, but instruments of their own upliftment. Critical to this is the knowledge by these peoples that they have a unique and valuable contribution to make to the advancement of human civilisation, that despite everything we have said, Africa has a strategic place in the global community.

In so doing he, presents himself, interestingly, as Nehru's heir. Here, India and Africa stand in a like relationship to the West—each with its own moral authority as it attempts to position itself to take on the task of modernization. Critically, however, Mbeki falters, much as Nehru did, as he seeks to explain how this moral authority might be translated into a manifesto for socialising the nation *through* its past and *into* an autonomous future. For a practical solution to *how* the authority of Africa is invoked Mbeki turns to the concept, more the imperative, of recreation and leisure, "without which" he says, "technological development will create the forces for its own destruction" (Mbeki, 2001:12). I am not going to rehearse the full argument here but the climax it reaches after sketching the full panorama of Africa's natural and human riches, suggests that "tourism should be treated as a critical corollary of modern scientific and technological development." The argument, of course, can be caricatured, but significantly, the point needs to be made, we are still left with the difficulty of working out the bases on which we frame our standards. As the debates in this conference suggest, we are not in agreement with each other, and even more disconcertingly, we are in something of a cul-de-sac in imagining how we might talk about the question of our standards, our own modernity.

In muddling our way through this situation, we, in the meanwhile, suffer within its grip. What we have in education in many countries shows the nature of this entrapment. In most countries, including our own, educational policy and development has come to settle around standards, call them normative markers if you like, of literacy and competence in the global economy. These advantage English-speaking middle-class groupings and disadvantage others who do not fit this profile or who struggle to obtain the attributes of English-speaking and middle-class behaviour. The result is that there is a disjunction between preferred pedagogical cultures almost everywhere and the daily lives that people live. The preferred form of school and the society in which it finds itself are in a state of non-alignment. Characteristic of these daily lives, it is argued, are dispositions and forms of deportment that are rooted in everyday knowledge (constituting what Muller (2000:79) would call the 'profane'), that have insufficient purchase in the world of the preferred school and what it deems to be good achievement. Confronting modern education in all of its complexity, young marginalized people, and sometimes their teachers, operate in a world signposted by people other than themselves. We have what Partha Chatterjee (1999:5) describes as "the imposition of high culture on society." The effect of this, as he says (Chatterjee, 1999:30), is that young people and their teachers have to work, with a "discourse… which, even as it challenged the colonial claim to political domination, it also accepted the intellectual premises of 'modernity' *on which colonial domination was based*."

This, I suggest, constitutes the central paradox for determining our own standards in many countries such as our own around the world. The paradox, for many amounts to the abidingly complex puzzle of how they might engage with this 'high' road and still remain alert to the challenges of including all their people and all their rich traditions and histories.

Even in countries such as Japan, Taiwan and South Korea, which stand as examples of alternative but successful forms of modernity, where it has been argued that Confucianism and systems of trust were powerfully influential in fostering economic development (see Amartya Sen, 2000), the paradox continues as Japan, for example, wrestles with the question of how to bring up its children and to deal with new social problems such as youth crime (see Numata, 2003: 261, and Kamijo, 2000:183).

What's missing? Another vision

What is missing in our educational systems is the lack of continuity between and amongst our major social institutions. There is a disjunction between the formality of high culture embodied in the school and the informality of everyday culture as lived in people's homes and families. This problem is of course universal, but it takes sharp expression here in Africa, and manifests itself as a poor alignment between school and home. Where fewest problems arise is when families are able to mimic the objectives and the interests of the school. Most problems arise when these are in weak articulation.

So what should we be doing? Should we be giving up our home cultures and taking on the cultures of the preferred form of school?

Towards answering the question let me quickly summarise the debate around globalisation and education. The nature of this debate has in recent years become a great deal more sophisticated. In its infancy the debate was essentially engaged as a conflict between dewy-eyed romantics seeing the dawn of a glorious age of anxiety-free social development, on the one hand, and dour pessimists who saw only the advance of a new rapacious imperialism on the other. We are now witnessing, in a range of fields of knowledge, much finer analyses of the significance of globalisation (see Dimmock and Walker, 2000; Nandy, 2000; and Held, 2003) that are alert to both its oppressive and liberatory potential. Similar possibilities, it is argued here, exist for discussions of globalisation and education.

These possibilities lie, I want to argue, in what already exists around us. I want to suggest that an alternative position to globalisation already exists in the practice of individuals and groups around the world that is premised on a critical synthesis of the insights of the Ghandians and the utopians, on one hand, and the Nehrus and the Mbekis, the new modernists on the other.

Central to the synthesis is what the utopians and the new modernists bring to a new and alternative pedagogical practice. Utopians bring a scepticism rooted in alternative understandings and knowledges of the world, while the new modernists bring a conviction that dominant forms of social practice, and their pedagogical forms, can be overhauled. It is in the interaction of these two core elements that new pedagogical practices can be envisaged that have the potential to resist those forms of globalisation that are only about standardisation, homogenisation and universalisation. How this interaction can be managed through educational practice is what we need to explore.

There are two directions in which this practice might go: a strong socially mobilised form and a weak state-driven programme of school change.

Strong forms of driving our modernity

In assessing possibilities for change, I suggest that critical engagements with globalisation *already* exist in practice in many societies and that instead of having to invent new ways of explaining how individuals and groups of people might mitigate, overcome and even transform the corrosive and exclusionary standardisation and homogenisation of globalisation, the challenge of contemporary education is how to make explicit pedagogical processes surrounding practices that already exist.

Significant in understanding this practice as it is *already-there* is its locatedness in what Gough and others, including Grossberg, have termed the third space or the gap. This third space is neither inside nor outside globalisation but pivots across the difference of between being inside and outside of it. Central about this third-space, it is argued here, is the fact that it already is a deep feature of every-day life in many societies. It is already present, for example, as young Chinese people encounter the similarities and differences of the old cultures their parents remember and cherish with the Western world presented to them in film, song, text and social practice. It is also present as young African people take the lessons of school with them into their initiation practices. Of profound importance is that we now have documented evidence of young people overturning the essence of their initiation into adulthood through the transformation of initiation from a physical encounter with hardship to a mental challenge of acuity. Less dramatically, and demonstrating the reverse direction of cultural flow, there is the mobilization of tradition in managing the everyday in the creation of social networks which not only help people survive within modernity but, significantly, to thrive. We have locally what Sen (2000) saw being used to great positive effect in places such as Japan with the recruiting of ancient cultural practices to help people establish very different corporate cultures in the workplace. Important about these examples is that they signify and carry embodied forms of learning. There is intense learning taking place as people effect these translations between their different cultural universes and create, for themselves, new, call it hybrid if you like, social cultures which are deeply rooted in their everyday worlds. This learning is implicit and informal, and so is simultaneously intensely straightforward but deeply complex. Important about it, is that it manages to work across the epistemological frameworks of the different worlds it inhabits.

A strong version of this third way seeks to understand the pedagogical character of this pivotal point and to see how it might be made explicit and so, therefore, inform our practice as educational people. I have elsewhere described the difficulty of attempting to put into words the nature of this translation and suggested that it amounts to what I called a crisis of representation. This crisis, I said, is a profound pedagogical moment, because one comes face-to-face with the possibility that one does not immediately have the words to capture meaning. The importance of this recognition is great. At its most complex, but also its most powerful, this is the critical moment in any learning process. It demonstrates to the learner and the teacher the hubris of any form of totalisation. Knowledge is never capable of being faithfully and completely reproduced. It is always in a state of interpretation. Looking at it in this way, all knowledge is therefore provisional and vulnerable. No knowledge is absolute. Globalisation as universalism, homogeneity and standardization as a new form of totalisation are therefore problematic. Knowledge itself must therefore be made the object of inspection

and not simply accepted or rejected because of where it comes from. What is its history, its objective and its scope?

But how do you codify what I have just described as pedagogy? In a very limited way, one can say, as Gough (2000) does, that it is a moment involving performativity. He says, "if knowledge is recognised as both representational and performative it will be possible to create a space in which knowledge traditions can be performed together" – a third interstitial space between the local and the global. This performance has people remaking meaning on their own terms.

How we turn this into practice is more easily said than done. It is at this point, however, that what is at stake in the struggle for a different kind of education becomes clearer. To the standardised forms that are taking root everywhere, minimally, we ought to be insisting on the development of deep forms of engagement with the *range* of knowledges that people have access to. Said (2002:46), talks of the "critical sense that can only come from a sustained encounter with the actualities of reading and interpretation." This reading, I want to suggest, need not only be textual. It must be that largely, but not only. It needs to include the oral, the visual and other forms of knowledge acquired in other kinds of ways. While this deep form of engagement does not guarantee that the young people who emerge from an encounter with it are, of necessity, better people, the chances are good that they will be able to deal with the full range of histories which are theirs and make their own form of modernity. The standards that they will commit themselves to in this version of modernity will be theirs which they would have constructed deliberately and self-consciously. If we can manage to codify this practice, we would be in a position of effecting a much closer correspondence between home and everyday on the one hand and school on the other. Approaching this we would have within our grasp the possibility of an African school, an African way to modernity – our modernity with our standards and our own benchmarks.

Weak forms of driving our modernity

The challenge of the strong form is that it demands a level of social dialogue which few societies are able to sustain. It demands also the commitment of people to understand themselves critically. It takes a commitment for the educational community, particularly teachers, to considerably raise the level of their game. It demands a commitment from parents to enter zones of public discourse that they would not normally recognize as their own. And critically, it calls on the state to facilitate the establishment of democratic arenas in which it itself participates as a servant of and a vehicle for the will of the people.

This is a hard task and so will be difficult to effect. In place of a socially driven dialogue about the nature of society, governments will have to take on the responsibility of interpreting the mandate given to them by their people. They will have to hold the debate about the nature of the modernity that people want. In leaving this responsibility at this level one has to recognise, however, that governments do not have the latitude of society and are less able to conduct themselves independently of the global order in which they find themselves.

This leaves one in the weak road. In this road, governments, in interpreting what is in the public interest will do what is manageable and pragmatic. In the absence of clear mandates from the people they will look for mandates elsewhere which have credibility. Significant guidelines they can draw on include those of UNESCO and other multinational declarations.

According to UNESCO (2002), regardless of gender, wealth, location, language or ethnic origin, quality education requires the following:

1. healthy, well nourished and motivated students;
2. well trained teachers and active learning techniques;
3. adequate facilities and learning materials;
4. a relevant curriculum that can be taught and learned in a local language and builds upon the knowledge and experience of teachers and learners;
5. respect for and engagement with local communities and cultures.
6. clear definition and accurate assessment of learning outcomes, including knowledge, skills, attitudes and values;
7. participatory governance and management; and
8. an environment that not only encourages learning but is welcoming, healthy, safe and gender sensitive.

These guidelines are important because they stress all the critical issues that are required for realizing a strong vision. But because they are driven by the state, their chances of being adopted in an organic way are reduced.

The state in this road has to lead by example. It has to take on the responsibility of winning people over to its interpretation of what is in their best interests. In the process it has to declare what the appropriate standards should be that it desires for itself and for the people within its authority. The political significance of this is great. In taking this political responsibility, it has to imagine itself into the world of the people and to think critically about how school and the everyday might be brought into a state of alignment. It has to take seriously the UNESCO injunctions of achieving the development of a well-trained core of teachers who can mediate a relevant curriculum "that can be taught and learned in a local language and build upon the knowledge and experience of teachers and learners". This is at the heart of what is required and demands that the state become a much more pedagogically minded state. In becoming this pedagogically minded state it has to do more than we are currently seeing in many countries, including South Africa, where the school is being held to standards which the system is not *geared* to achieving. It is not that these standards are wrong but what we are seeing here is the state thinking *against* local conditions and the local environment as opposed to *from within* these local conditions. While elements of national systems are attuned to the interests and investments of local conditions, such as components of the new curriculum in South Africa, they are only add-ons. Thinking in relation to where the people are demands a much deeper understanding of how to engage with the complexity of the local.

References

- Adams, R. (1997) Epistemology, Afrocentricity, and Ideology. In P. Logan (Ed). *A Howard Reader: An Intellectual and Cultural Quilt of the African-American Experience.* Boston, MA: Houghton Mifflin Custom Publishing.
- Banteyerga, H. (1994) An Alternative Model in Teacher Education: The Classroom in Focus. A Paper presented at the Pan-African Colloquium: Educational Innovation in Post-colonial Africa. University of Cape Town, Rondebosch.
- Bartolovich, C. (2002) Introduction: Marxism, Modernity and Postcolonial Studies. In C. Bartolovich and N. Lazarus (Eds). *Marxism, Modernity and Postcolonial Studies*. Cambridge: Cambridge University Press.
- Brady, W. (1997) Indigenous Australian Education and Globalization. In V. Masemann and A. Welch (Eds). *Tradition, Modernity and Post-Modernity in Comparative Education.* Dordrecht: Kluwer Academic Publishers.
- Castells, M. (1996) The Information Age: Economy, Society and Culture, Vol 1: The Rise of the Network Society. Oxford: Blackwell Publishers.
- Chatterjee, P. (1997) 'Our Modernity', a lecture published by the South-South Exchange Programme for Research on the History of development (SEPHIS) and the Council for the development of Social Science Research in Africa, Rotterdam/Dakar, 1997.
- Chatterjee, P. (1998) 'Beyond the Nation? Or Within?' In *Social Text,* 16 (3), pp. 57-69.
- Giddens, A. (1991) Modernity and Self-Identity: Self and Society in the Late Modern Age. Stanford, CA: Stanford University Press.
- Gough, N. (2000) Globalization and Curriculum Enquiry: Locating, Representing and Performing a Transnational Imaginary. In N. Stromquist and K. Monkman (eds). *Globalization and Education: Integration and Contestation Across Cultures.* Lanham, Maryland: Rowman and Littlefield.
- Grossberg, L. (1994) Introduction: Bringin' It All Back Home – Pedagogy and Cultural Studies. In H. Giroux and P. McLaren (Eds). *Between Borders: Pedagogy and the Politics of Cultural Studies.* New York: Routledge.
- Kamijo, M. (2000) Education in the 'Melting Society': New Challenges Facing Japanese Schools and Society. In K. Mazurek, M. Winzer and C. Majorek (Eds). *Education in a Global Society: A Comparative Perspective.* Boston: Allyn and Bacon.
- Nandy, A. (2002) Time Warps: The Insistent Politics of Silent and Evasive Pasts. Delhi: Permanent Black.
- Nekhwevha, F. (1999) No Matter How Long the Night, the Day is Sure to Come: Culture and Educational Transformation in Post-Colonial Namibia and Post-Apartheid South Africa. In C. Soudien and P. Kallaway (Eds). *Education, Equity and Transformation.* Dordrecht: Kluwer Academic Publishers.
- Numata, H. (2003) What Children Have Lost by the Modernisation of Education: A Comparison of Experiences in Western Europe and eastern Asia. In M. Bray (ed). *Comparative Education: Continuing Traditions, New Challenges, and New Paradigms.* Dordrecht: Kluwer Academic Publishers.
- Prakash, G. (1999) *Another Reason.* Princeton: Princeton University Press.

- Said, E. (2002) The Book, Critical Performance and Education's Future. In K. Asmal and W. James (Eds). *Spirit of the Nation: Reflections on South Africa's Educational Ethos.* Claremont, Cape Town: New Africa Education.
- Sen, A. (2000) "Culture and Development". Talk at World Bank Meeting, 13 Dec 2000. Mimeo.
- Soudien, C. (2004) Inside but Below: The Puzzle of Education in the Global Order. In J. Zajda and R. Zajda (Eds). Globalisation, Education and Policy Research: Changing Paradigms. International Handbook on Globalisation, Education and Policy Research. Kluwer Press.
- UNESCO (2002) Report of the Conference of Ministers of Education of African Members States (MINEDAF) Dar-es-Salaam.
- Wright, H. (2002) 'Editorial: Notes on the (Im)Possibility of Articulating Continental African Identity', in *Critical Arts,* 16 (2), pp. 1-18.

Accountability and support in school development in South Africa

Nick Taylor, JET Education Services

Abstract

South Africa has the one of the highest rates of inequality in the world, and certainly the highest in Africa by a large margin. Increasing the quality of schooling provides poor children with the best chance of emerging from poverty. Yet South African children score significantly lower on cross country assessment exercises than many of their much poorer neighbours.

School improvement has been an area of intense activity by donors and NGOs in South Africa for over two decades. Government has also entered this terrain in the last five years. Although research on school improvement is a more recent development, several studies have begun to identify important lessons, both for the design of intervention programmes and for further research.

Well established findings from other countries are confirmed by the South African research. Thus poverty and home factors emerge as the most powerful determinants of educational opportunity. Details are described about home educational practices. At the school level, time management and instructional leadership are the strongest factors associated with enhanced learning. Classroom level indicators remain elusive, probably because of the paucity of longitudinal designs. Nevertheless, descriptions of teaching practices are indicative of very low levels of reading and writing in the majority of South African schools.

A number of research questions remain unanswered, while the findings that have been established require replication. The paper ends with proposals for research programmes designed to extend the knowledge base on schooling.

Introduction

Around the world schools are under two kinds of social demands: the press for increased access, and the demand for graduates with ever more sophisticated cognitive skills. These pressures are frequently in conflict with one another. South Africa is no exception and, while we have done relatively well in providing school places to close to 100% of children of primary school age and an increasing proportion of children of high school age, there is also evidence that this press for equity has come at the expense of quality (Crouch and Fasih, 2006). As the South African economy begins to grow at levels last seen decades ago, it is clear that a lack of skills is exerting a brake on both national development and the social mobility of poor South Africans. The quality of schooling is key to the quality of skills in every sphere of social and economic life. This paper examines recent efforts in South Africa to improve school quality through the twin tools of accountability and support.

Standards-based accountability

Standards-based accountability[1] (SBA) has assumed increasing importance in many countries over the last two decades or more, as a means of improving school outcomes. According to Carnoy et al (2003), by 2001 49 states in the US had adopted some form or other of this approach. The assumptions underlying standards-based accountability are:

- clearly defined standards, in the form of a common curriculum, explicate to teachers and pupils what is to be learnt;
- state-wide or national tests assess the extent to which schools and pupils are achieving the standards;
- rewards and sanctions accompany the results of the tests: for pupils these occur in the form of access to further and higher education and the job market, thus having an important bearing on career choices; for schools the consequences come in the form of market information, signaling the quality of the school to parents, or via direct rewards and sanctions administered by the district, state or national bureaucracy.

In the face of its popularity with education administrators, SBA has generated strong criticism from academic circles, where four arguments against this approach are raised. First, it is argued, the high-stakes nature of the testing component leads to distortions in the curriculum. The main kind of distortion arises from the form which large-scale tests sometimes adopt (multiple-choice, solving problems of a restricted nature, individual responses), leading to a focus on certain parts of the curriculum and a neglect of others (extended writing, solving complex problems, teamwork). This narrowing of the curriculum is further exacerbated when testing is restricted to certain subjects or certain grade levels of the school, and when teachers 'teach to the test', drilling pupils on the kind of items commonly encountered in the tests, at the expense of other parts of the curriculum.

A second common criticism of SBA is that it takes no account of the value which a school adds. This line of argument arises from the strong correlation between educational outcomes and socio-economic status (SES). For example, a school situated in an affluent area and attended by pupils from middle class homes may be underperforming relative to its SES

[1] Also known as performance-based reform (Hopkins, 2002) or outside-in reform (Muller, 2000).

peers, but easily outscores a poorer school which is doing a really good job with children from working class homes. The latter school would clearly be at a disadvantage if judged on a common 'league table' in which no account is taken of SES, while the underperforming high-SES school would look good.

A third criticism leveled against the standards-based approach is that it generates high levels of anxiety among children, for whom success or failure in one set of tests at one point in their lives will have a profound influence on the rest of their careers.

Carnoy et al (2003a; Elmore, 2003) raise a fourth problem with accountability systems linked to high-stakes tests, which is perhaps more fundamental than the previous three lines of argument. These authors conclude that such systems are effective in raising aggregate student scores, but that there is wide variability among schools in their specific responses to standards-based reform initiatives. They argue that the response of any particular school to *external accountability measures* depends on the state of its *internal accountability systems*. The latter are defined as the collective expectations held by members of the school community, together with the organizational rules, incentives and implementation mechanisms that constitute the school's formal accountability system (including instructional practices and supervision processes).

Strong internal accountability systems enable a school to respond positively to standards-based reform programmes, but schools whose internal systems are weak are unable to respond. The majority of schools, according to Carnoy et al (op cit), fall into the latter category, and the key to improving their ability to respond is capacity building, aimed at aligning and strengthening internal accountability systems.

These ideas frame our examination of the evidence for the effects of SBA and capacity building (support), in various combinations, on school performance in South Africa over the last decade.

SBA can have significant positive effects on targeted indicators

A feature of the post-1994 profile of SC results was a steady decline in both the pass rate and the proportion of pupils attaining university exemption. While the number of candidates fluctuated between 450 000 and 550 000 over the period 1994 to 1999, the pass rate declined from the 58% to 49%, and the exemption rate dropped from 18% to 12% (see Table 1). This decline should not be surprising, given the thorough-going reorganization of the entire school system following the change in government in 1994 and the consequent destabilization this must have had on schools. However, after the second general election of 1999 government began to pay serious attention to the question of Senior Certificate (SC) results, the only quality indicators available, then and now. Among a number of measures directed at improving the functionality of schools and the implementation of the new curriculum and other policies, the national DoE established a National Monitoring Forum, the aim of which was to co-ordinate improvement in the SC exam results (MoE, 2000). Each province was required to institute a SC improvement plan with a special focus on underperforming schools, defined as those which achieved pass rates in the 0-20% category.

The results of these efforts was immediate, with pass and exemption rates showing a dramatic turnaround in 2000 and a sharp upward trajectory in subsequent years.

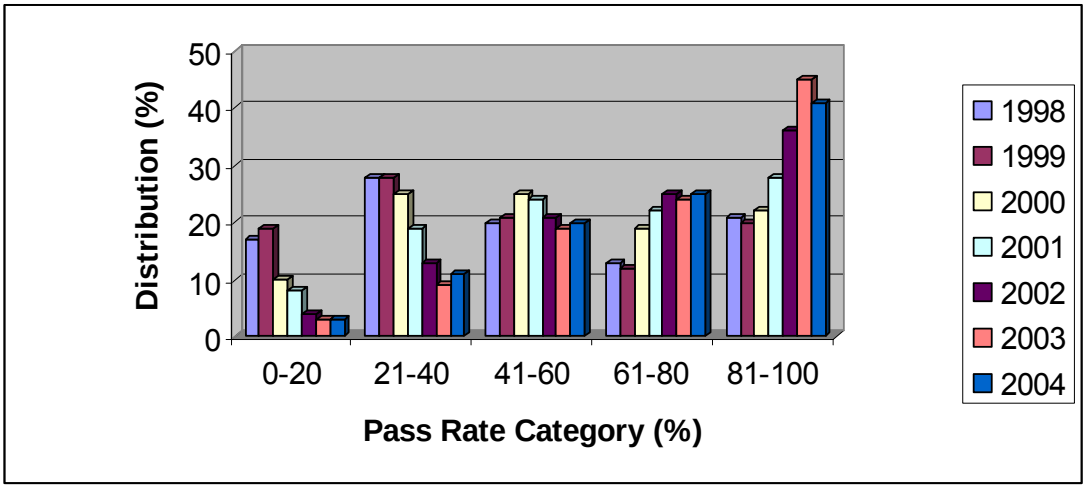
Table 1: Senior certificate examination results, 1994-2002

	Candidates	Total Passes	%	University Exemption	%
1994	495 408	287 343	58	88 497	18
1995	531 453	283 742	53	78 821	15
1996	518 032	278 958	54	79 768	15
1997	555 267	261 400	47	69 007	12
1998	552 384	272 488	49	69 856	13
1999	511 159	249 831	49	63 725	12
2000	489 941	283 294	58	68 626	14
2001	449 371	277 206	62	67 707	15
2002	471 309	324 752	69	75 048	16
2003	440 267	322 492	73	82 010	19
2004	467 985	330 717	71	85 117	18
2005					

Source: DoE, 2004, 2005a, 2005b

As we shall show later, the effects of these SBA measures were not felt equally in all schools. Nevertheless, they did effect schools in all pass rate categories, with the numbers of schools achieving 0-20% and 21-40% dropping sharply, and the numbers in the two highest deciles increasing, and the middle decile remaining more or less constant (see Figure 1).

Figure 1: Percentage distribution of schools by pass rate category, 1998-2004



Source: Collated from DoE, 2004, 2005a, 2005b.

The most intense SBA initiative at the provincial level was the Education Action Zone (EAZ) programme adopted by the Gauteng Department of Education (GDE) in 2000. Seventy schools in the province which exhibited pass rates below 20% were targeted for a package of interventions. The EAZ was designed as a comprehensive systemic initiative which would include monitoring schools and providing support and training to principals,

teachers and pupils. However, in reality, the programme did not fully meet its systemic intentions, focusing largely on accountability measures (Fleisch, 2001; 2003; 2006). A project approach was adopted in administering the programme, rather than strengthening the systems and capacity for school monitoring and support in the standard line functions of the GDE: thus, the EAZ was managed from the provincial head office, with special units responsible for earmarked schools and reporting directly to the provincial Minister of Education and the Head of Department.

The EAZ achieved an impressive rise in SC results in targeted schools on a range of indicators: numbers of candidates passing at both higher and standard grades (HG, SG), overall pass rate, university exemption rate, and the numbers of A symbols achieved by pupils (80% or more on aggregate across all subjects) (Table 2).

Table 2: SC Results in Education Action Zone Schools, 1996-2003

	1996	1997	1998	1999	2000	2001	2002	2003
Pass Rate (%)	**24.32**	**17.16**	**20.49**	**15.59**	**33.09**	**48.59**	**59.19**	**66.48**
University Pass Rate (%)	**2.25**	**1.76**	**1.62**	**0.99**	**2.10**	**3.53**	**4.59**	**6.52**
Number Passed Higher Grade	**211**	**167**	**155**	**104**	**170**	**197**	**246**	**387**
Number Passed Standard Grade	**2015**	**1455**	**1467**	**1530**	**2508**	**2515**	**2024**	**3631**
Total Passed	**2226**	**1622**	**1622**	**1634**	**2678**	**2712**	**3170**	**4018**
Number A Symbols Awarded	**8**	**3**	**28**	**32**	**200**	**124**	**216**	**373**
Number Wrote Higher Grade	**8455**	**7921**	**6254**	**4252**	**1778**	**1526**	**1383**	**1571**
Number Wrote Standard Grade	**942**	**1551**	**3334**	**6229**	**6315**	**4055**	**3973**	**4468**
Total Wrote	**9397**	**9472**	**9588**	**10481**	**8093**	**5581**	**5356**	**6039**

Source: Fleisch, forthcoming.

Not only are these results very impressive on their own, but EAZ schools also increased relative to other schools in the province: thus, in the first two years the aggregate pass rate for project schools increased by an average of 14,5%, which exceeded the improvements shown by both other former DET schools in the province (up 10,1%), and all public schools in Gauteng (5,3%) (Fleisch, 2001).

The evidence presented this far clearly shows that the pressure created by SBA measures can result in dramatic increases in targeted indicators on a large scale. The fact that the improvements shown by EAZ schools exceeded that of comparable sub-categories of schools in the province, suggests that the greater the pressure exerted by SBA, the greater the gains. Whether these indicators measure anything worthwhile, and whether increases in the indicators reflect improved performance are questions to which we now turn.

SBA may lead to curricular and other distortions

Skepticism has been expressed about the meaning of the changes exhibited both by the national trends shown in Table 1 and the EAZ results in Table 2. Indeed, in the very first year of the turnaround, Professor Jansen expressed doubt as to whether they reflected anything more than 'imaginative statistics' (Fast Facts, 2001, 3). Similarly, while conceding that improvements in the results of EAZ schools were spectacular, Fleisch (forthcoming) argues that, instead of reflecting real increases in quality, these changes may have been achieved through manipulations of a more cynical nature. Fleisch advances four kinds of

alternative explanations: reducing the number of SC candidates, through the exclusion of high-risk pupils; encouraging pupils to write papers at the easier standard grade level; lowering the standard of examination papers; or changes in the moderation process. We examine each of these alternative explanations in turn.

Excluding high-risk candidates

There is no doubt that the improvements in SC results over the period 2000-2003 were accompanied by significant reductions in the numbers of pupils registered for the exam in EAZ schools. The number of candidates dropped by 4 442 (42%) between 1999 and 2003. However, in the face of a concomitant rise of 2 384 (146%) in the absolute number of pupils passing, it seems obvious that clearing grade 12 classes of pupils who were unlikely to pass, allowed greater attention to be focused on more promising candidates, thus achieving not only very significant quantitative gains in educational opportunity for these pupils, but also doing so at vastly improved rates of efficiency.

Increasing the ratio of SG:HG passes

In querying the value of the improvements shown by EAZ schools, the second kind of argument raised by Fleisch is that the increases in pass rate may have been achieved by shifting the pattern of registration away from the HG level towards SG, thus making it easier to pass. As Table 3 shows, the data does not support this hypothesis.

Table 3: Pass rates compared with SG:HG ratios, EAZ schools, 1996-2003

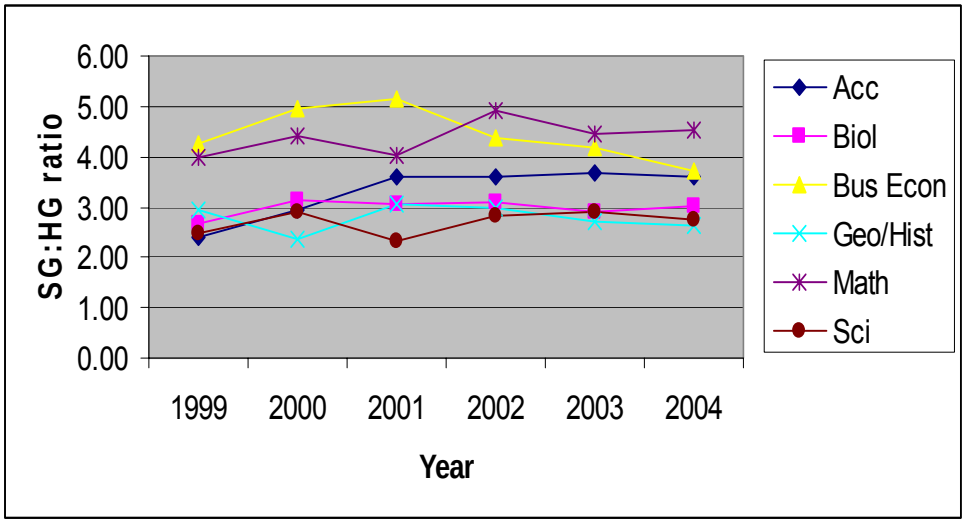
	1996	1997	1998	1999	2000	2001	2002	2003
Pass Rate (%)	**24.32**	**17.16**	**20.49**	**15.59**	**33.09**	**48.59**	**59.19**	**66.48**
SG:HG	**9.55**	**8.71**	**9.46**	**14.71**	**14.75**	**12.77**	**8.23**	**9.38**

Source: Calculated from Fleisch, forthcoming.

In the EAZ sample in Gauteng, a sharp rise in the SG:HG ratio from 1998 to 1999 was accompanied by a fall in pass rate; conversely, over the period 2000-2003 a steady rise in the pass rate was accompanied by falling SG:HG ratios.

Nor is there a clear association between increasing pass rates and rising SG:HG ratios in the selected subjects for which the DoE publishes national figures, as shown in Figure 2.

Figure 2: SG:HG ratios for selected subjects, national means, 1999-2004



Source: Calculated from DoE, 2004, 2005a, 2005b. Figures for Geography only published for 1999 and 2000; thereafter History is shown instead.

Against steadily rising pass rates, the SG:HG ratio for 5 of the 6 subjects did rise from 1999 to 2000, but after this only Accountancy continued to rise, while the ratios for three other subjects fell in 2001 while a fourth declined a year later. After this the patterns tend to flatten or decline slowly. The available evidence indicates that there is no systematic relationship between SG:HG ratios and rising pass rates.

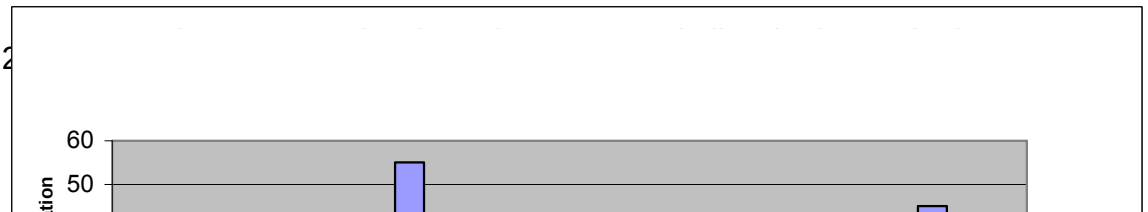
Lowering the standard of the exam papers

In response to growing public concern that the pass rate improvements since 1999 were the result of lowered standards, Umalusi commissioned a research project to investigate the issue. Question papers from 1992 to 2003 in Mathematics, Physical Science, Biology, History, English Second / Additional Language and English First Language were evaluated. A three-member team of experts was appointed for each of the six subjects to determine whether a discernable drop in standards had occurred over this period.

One of the specific questions posed to subject teams was whether the proportion of items representing different levels of cognitive challenge had changed. For the purposes of this exercise, individual items were rated on the scale of 1 – 3, with (1) indicating the simplest and most basic question types and/or knowledge required to answer them, (2) indicating questions of average difficulty, and (3) questions involving more sophisticated linguistic, literary and general knowledge.

A finding which was common to the report of most of the teams was that the standard varied across the different papers set by various examining authorities over the period in question. This is well illustrated in the graph provided by the English Second/Additional Language team (ESL).

Figure 3: Comprehension and Language: mark allocation by authority and year.



Source: Yeld et al, 2004

A second point made very strikingly by the data shown in Figure 3 is the change in the proportions of item types between the 2001 and 2003 national ESL Language papers, with a rise in the number of easier questions (Level 1) and a fall in the number of questions requiring higher levels of understanding (Level 3). In conclusion, the team noted that 'From the analysis of the papers above, it is difficult to escape the conclusion that the nationally set paper (Paper 1) is becoming easier—or, in the jargon of the examiners, becoming 'more accessible" (Yeld et al, 2004, 14).

'Accessibility' had clearly become the watchword for examiners in a number of subjects, and the conclusion reached by the ESL team was confirmed, in one way or another, by all but one of the other subject teams. The Biology team noted a trend toward setting a larger number of questions, each containing fewer marks, and that a consequence of this practice is that '… learners are seldom asked to demonstrate in-depth knowledge of important biological processes…', that 'The many short questions permit a fairly superficial coverage of all topics, but do not probe learners' ability to articulate their answers in a written form, nor to illustrate their answers appropriately', and that '…opportunities to test ability to synthesize information from different topics are limited' (Dempster et al, 2004, 14).

Although they had access to a wide variety of papers from different authorities, because of the limited number of papers which provided valid comparison points across time, the English First Language team decided to compare only the 1999 and 2003 papers set by the Gauteng Education Department. The team found an overall decline in cognitive challenge in the language paper, and to a lesser extent in literature. They noted that in 2003 very little reading or specific knowledge of English was required for language, while for literature many questions could be answered without knowledge of the full texts of a number of set works (Allais et al, 2004).

The History team noted with approval the approach adopted by the new curriculum, but concluded that, although assessors paid lip service to these principles, ' … it in fact did little more than reward rote learning under a different guise', and that ' … there is clearly less emphasis on critical thinking and more reliance on questions which identify tasks in terms of

"describe" or "discuss" rather than in terms of "explain" or "analyze" ' (Kallaway et al, 2004, 10). As a consequence the team expressed strong feelings that ' … these phenomena had a considerable effect on the profile of marks received in 2003 and boosted the marks unjustifiably' (ibid, 12).

In Mathematics the SG and HG papers for the 1999 and 2003 national examination were compared. At SG, for both papers 1 and 2, the team described a 'dramatic' decrease in the number of Level 3 questions in comparison to Level 1 items (Carrim et al, 2004). The same comparison was not done for the HG papers, although the report noted that in the national Paper 2 only 3% of marks were allocated to Level 3, compared with 25.5% in the corresponding IEB paper.

Science was the only subject in which the standard, as a whole, was found to be consistent over time: while the 2003 Physics paper was slightly easier than previous papers, the Chemistry paper was slightly more difficult (Masemula et al, 2004).

Changes in the moderating process

In addition to assessing the quality of exam papers, the Umalusi review also looked at the moderation of raw scores. The latter review noted that, because of the large differences in standard across pre-1994 examining authorities, many students were at a disadvantage in a common system. To compensate, adjustments performed at statistical moderation meetings from 1996 to 2001 were generally upward, frequently by the maximum of 10% allowed. Furthermore, ' … [o]ne of the consequences of the predominant trend of upward adjustments was a reluctance by the new examination authorities to accept downward adjustments when these were recommended by the SAFCERT / UMALUSI statistics team. … This certainly has resulted in an upward movement in pass rates' (Umalusi, 2004, 8).

Conclusion

What are we to conclude from the above evidence? Is it clear that the large rises in our indicators for the SC exam results were largely due to easier papers and mark manipulation? Certainly, the trend turned as soon as the Umalusi research was made public, and the standard-setting body wrested control of the moderation process from the DoE in 2004 and exerted closer oversight of the standards of the exam papers set in 2005. These steps were accompanied by an immediate change in SC results, showing a downturn in 2004 which was continued into 2005 (see Table 1).

But were all the gains of 1999-2003 due to the manipulation of standards and scores, or was there also some improvement in delivery by schools, either in tighter management or improved teaching, or both? Evidence for the effects of SBA on school level improvements is less conclusive than the strong case we have presented so far for the manipulation of results. What is blindingly clear is that the primary purpose of SBA is defeated if standards are not kept constant or improved over time. It is also crystal clear in the South African case that high school standards are very low, focusing largely on cognitive skills of an elementary nature, at the expense of higher order processes such as analysis, synthesis, and the development of written argument. The best intentions of SBA are to improve standards, but,

if close watch is not kept on all the exits, then it can have the opposite effect, and our analysis so far indicates that the DoE succumbed to this temptation in the years 1999-2003.

The good news is that Umalusi stopped the rot, following its legislative brief to be the watchdog of standards. Whether it manages to improve the quality of education, in the face of all contrary temptations, will be a matter for history to judge. Umalusi's first task in improving the quality of schooling is to steadily improve the standards of the exam papers, accompanied by an explicit description of how this is to be done so that teacher and pupils are able to prepare for the new demands.

By far the most important step towards improving standards is to undertake a radical overhaul of the ESL curriculum. Over 80% of SC candidates not only write this subject, but use the conceptual and language skills it provides to learn all their other subjects. Yet, the findings of the Umalusi research indicate that the current low standards of ESL are severely aggravating the educational disadvantages which ESL speakers already labour under because of poverty. Furthermore, the practice of compensating these candidates by adding 5% to all their non-language scores, instituted in 1998 as a temporary measure, is counterproductive to improving the quality of schooling. These are sensitive issues and, while taking the current easy route may bring short-term political gains, the long-term educational damage accretes year by year.

SBA and support together leverage significant improvements in school performance

The Quality Learning Project (QLP) was programme which provided support measures to schools, intended to complement the SBA effects of the SC regime in South Africa. Working in 524 high schools across the nine provinces provided training and support to teachers, principals and district officials. A longitudinal evaluation (HSRC, 2005) found that over the life of the project QLP schools achieved significantly better results in the SC exam than selected control schools, in terms of greater numbers of overall passes, university exemptions, passes in mathematics, and in the pass rate. Similarly, improvements shown by QLP schools were significantly better than those of the national mean (Table 4).

Table 4: Comparison of QLP SC results with the national mean, 2000-04

	Increase 2000 – 2004								
	Passes		Exemptions		HG maths		SG maths		% Pass
	No	%	No	%	No	%	No	%	Change
Total QLP	4167	18.3	1182	34.8	585	152.3	8741	137.5	14.0
Total SA	47314	16.7	16493	24.0	8466	47.0	46512	58.0	12.8
Difference*		1.6		10.8		105.0		79.0	1.2

*Computed by subtracting the improvements exhibited by the national mean over the life of the project from those exhibited by QLP schools.

Source: Taylor and Prinsloo, 2005.

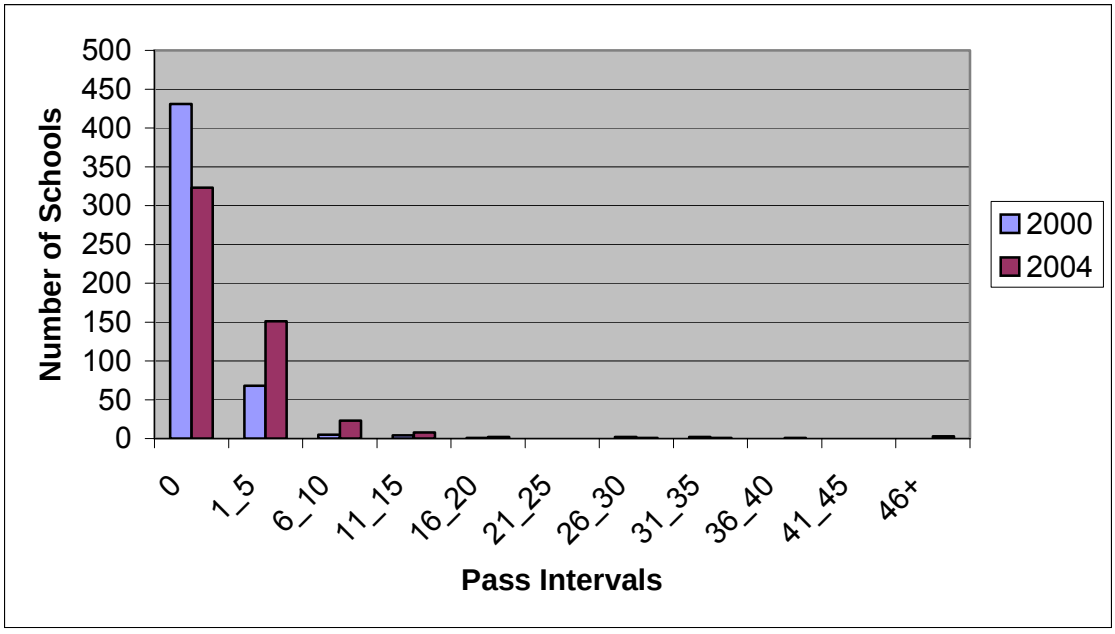
The evaluation concluded that the project had a significant effect on the performance of these very poor, largely rural schools: the QLP support programme resulted in improved practices in school management and classroom teaching, which in turn produced improved learning (HSRC, 2005). In Carnoy et al's (2003) terms, the support provided by the QLP assisted schools to strengthen their internal accountability systems, which in turn enabled them to meet the external accountability challenge posed by the intense SBA expectations of the time. Furthermore, we infer that the additional gains made by QLP schools, in comparison to the national mean, resulted from improved learning rather than any manipulation of the results, since the latter factor was constant for all schools.

Interestingly, the QLP evaluation also tracked pupil performance on math and language tests at grades 9 and 11. The only learning gains discernible were improvement in writing at grade 11. The most likely explanation for this disappointing result, in the light of the very impressive improvements at SC level, is that, whereas intense pressure is put on schools to perform in the SC exams, no such pressure is applied at lower levels of the system. Hence, capacity-building is most effective when accompanied by SBA measures.

SBA and support have no effect on some schools

Impressive as the QLP results shown in Table 4 are, they also have a dark side. This is shown in Figure 4, a frequency distribution of QLP schools, by number of HG maths passes achieved, before and after the intervention.

Figure 4: Comparison of HG maths passes in QLP schools, 2000 and 2004



These figures reveal that the improvement in maths results shown in Table 4 were achieved by only one-third of the schools. Fully two-thirds remained without a single HG maths pass at the end of the programme: these schools gained nothing, in terms of improving their HG Mathematics passes, from the considerable resources expended on them over a period of 5 years.

These schools would be classified as 'failing' in the three-part classification devised by Hopkins et al (Hopkins, Harris and Jackson, 1997). Rewards and sanctions have no effect in these situations, as the schools are unable to help themselves: they require a high level of external intervention and support. There should be a clear and concerted focus on a specific, limited number of factors. In many schools in this state the first thing to be done is to remove the principal; often strong mediation is required to break situations of conflict between various groups in the school. Only government has the authority to intervene here.

One of the most important lessons of the QLP is that programmes of this kind are only successful in schools which have a minimum level of capacity at the start. This is a lesson that government has itself learned from the Dinaledi project, which was in many ways similar to the QLP[2] and which also had little success in a significant number of target schools: thus, the second phase of Dinaledi is being targeted at moderately functioning schools.

Conclusion

Standards based accountability consists of setting clearly defined expectations for pupil learning, and exerting pressure on schools, teachers and pupils to meet these standards. Around the world SBA has become the dominant mechanism adopted by politicians and administrators for attempting to improve the quality of schooling. However, as is the case with most major policy initiatives in education, its effects are complex, mixed and often counter-intuitive, despite the best efforts of its proponents and detractors to paint it either as a simple solution to the problem of mediocre quality, or as the greatest enemy of progressive schooling. The South African evidence to date around SBA can be summarized in the form of five propositions.

Proposition 1

SBA, on its own, can lead to very significant changes in the indicators chosen to signal successful learning, and the higher the pressure on schools the greater the change.

What gets measured is what gets changed, and it is therefore important that the indicators are carefully chosen to reflect true measures of quality schooling. In South Africa there are two major problems with current standards. The first is that those defining the skills required by high school leavers are too low: this, together with poor school performance, is the root of the country's skill problem. Standards need to be moved progressively up the scale of cognitive demand, across the full range of school subjects, but most importantly for English Second Language.

The second problem with South Africa's standards is that they are defined, measured and valued at only one point in the system, the end of high school. Standards are best built

2. The Dinaledi project, implemented in 102 schools nationally, achieved similar results to those shown in Table 3. However, one important difference between the two programmes is that, although Dinaledi was also targeted at very poor schools, it started from a significantly higher base than did the QLP. Thus, the frequency distribution of HG SC maths passes at the start of Dinaledi was skewed to the right when compared with the QLP distribution shown in Figure 4. Nevertheless, a very significant proportion of Dinaledi schools also showed no improvement as a result of the intervention (Taylor and Prinsloo, 2005).

incrementally, and clear signals need to be conveyed to and their achievement measured at all levels of the system. In order to address this problem, government's Systemic Evaluation process, which measures pupil performance in Mathematics and language at the end of each school phase, needs to be improved and used for accountability and support purposes.

Proposition 2
There is a tendency to accommodate SBA pressure by the easiest means available, such as manipulation of the indicators, without necessarily improving learning.

The inclination to manipulation underlines the need to ensure that the intended standards are realized at all stages of the SBA process, including setting exam papers, marking pupil responses and moderating raw scores. Umalusi is in the process of limiting indicator manipulation in the first and third of these areas, but a major gap persists with respect to the marking of papers. This is a key node in the SBA process: it provides the opportunity to build capacity among teachers in the form of the fine-grained professional judgement required to recognize various manifestations of the standards; this is the first step towards replicating them in classrooms.

Proposition 3
There is insufficient evidence to indicate whether SBA pressure, on its own, leads to improved school management, classroom teaching and pupil learning.

Strong evidence of manipulation in the period 1999-2003 has raised questions about the validity of indicators used to measure school quality in South Africa. As a result, the question of whether school performance improved over this period is difficult to answer. Circumstantial evidence indicates that some management slack was squeezed out of the high school system, resulting in institutions which are at least more efficient in moving pupils along. However, it is unlikely that SBA has resulted in significantly better teaching, a task which will require considerable capacity building.

Proposition 4
A combination of SBA and capacity building aimed at strengthening internal accountability systems has significant effects on school performance.

Programme evaluations have begun to demonstrate that a number of school development projects in South Africa are achieving improved performance. Specific factors identified as key elements of the internal accountability systems required for effective teaching and learning include: time regulation, planning and monitoring curriculum coverage, and management of textbooks. However, this is work in progress and much ground needs to be covered before the complex processes of curriculum leadership and teaching are properly understood.

Proposition 5
In the poorest performing schools the combination of SBA and support has no effect.

The fundamental conditions conducive to effective learning do not exist in the majority of schools in South Africa. Such schools are impervious to any combination of SBA pressure and capacity building tried thus far. These schools need something else, and the international literature indicates that what is required is organizational development: fundamental issues such as removing ineffective school principals[3], mediating conflict, and building administrative capacity require attention before any learning is possible. Government has signaled its intention to move into this area, but this is an enormous task, given the weak state of the educational bureaucracy at all levels of the system.

References

Allais, J., Davidson, E, and Reed, Y. (2004) Report on English Senior Certificate Examinations. Pretoria: Umalusi.

Carnoy, M. Elmore, R. and Siskin, L. (2003a) Introduction, in M Carnoy, R Elmore and L Siskin (eds) *The New Accountability: High Schools and High-Stakes Testing*. New York: RoutledgeFalmer.

Carnoy, M. Elmore, R. and Siskin, L. (2003b) (eds) *The New Accountability: High Schools and High-Stakes Testing*. New York: RoutledgeFalmer.

Carrim, I, Julie, C., and Kitto, A. (2004) Report on an Investigation into Senior Certificate Mathematics Examinations of Various Years offered by the National Education Department and the Independent Examination Board. Pretoria: Umalusi.

Crouch, L. (2005) 'South Africa and quality reforms: possible lessons'. In *Journal of Education for* International Development, 1(1).

Crouch, L. and Fasih, T. (2006) Patterns in Educational Development: Implications for Further Efficiency Analysis. Mimeo.

Dempster, E. Khumalo, G. and Mavovana, M. (2004) Umalusi Senior Certificate project: Biology group report. Pretoria: Umalusi.

DoE (2004) Education Statistics in South Africa at a Glance 2002. Pretoria: Department of Education.

DoE (2005a) Education Statistics in South Africa at a Glance 2003. Pretoria: Department of Education.

DoE (2005b) Education Statistics in South Africa at a Glance 2004. Pretoria: Department of Education.

Elmore, R. (2003) Accountability and Capacity, in M Carnoy, R Elmore and L Siskin (eds) *The New Accountability: High Schools and High-Stakes Testing*. New York: RoutledgeFalmer.

Fast Facts (2001) 'Hold your applause for education'. In *Fast Facts*, February 2001. Johannesburg: South African Institute of Race Relations.

Fleisch, B. (forthcoming) 'Bureaucratic Accountability in the Education Action Zones of South Africa'. To appear in *South African Journal of Education*.

Fleisch, B. (2001) Draft Research Report on the Education Action Zones, GDE. Joint Education Trust: mimeo.

[3] In February 2006 both the national Minister and the Gauteng MEC for education threatened to close dysfunctional schools, but after protests from students, nothing has been heard of this issue since.

Hopkins, D. (2002) The Aga Khan Foundation School Improvement Initiative: An International Change Perspective, in Anderson, SE (ed) *Improving Schools Through Teacher Development: Case Studies of the Aga Khan Foundation Projects in East Africa*. Lisse, Netherlands: Swets and Zeitlinger.

Hopkins, D., Harris, A., and Jackson, D. (1997). 'Understanding the school's capacity for development: growth states and strategies'. In *School Leadership and Management*, 17(3),401-411.

HSRC (2005) Improving Learning in South African Schools: Executive Summative Report to the Business Trust on the Quality Learning Project (QLP). Pretoria: Human Sciences Research Council.

Kallaway, P. Alexander, M. and Angier, K. (2004) Senior certificate research evaluation exercise: History report. Pretoria: Umalusi.

Koretz, D. (2005) 'Alignment, high stakes and the inflation of test scores'. In *Uses and misuses of data for educational accountability*. National Society for the Study of Education: Yearbook 2005.

Masemula, M., Karsan, P., and Clarke, J. (2004) An Investigation into the Senior Certificate Physical Science Examinations (1992-2003). Pretoria: Umalusi.

MoE (2001) Fourth Report to the President from the Minister of Education. Pretoria: Ministry of Education.

Nicols, S. and Berliner, D. (2005) *The inevitable corruption of indicators and educators through high stakes testing.* Arizona State University: Education Policy Studies Laboratory.

Muller, J. (2000) The sound and fury of international school reform: A critical review. Joint Education Trust, mimeo.

Taylor, N., and Prinsloo, C. (2005) The Quality Learning Project: Lessons for High School Improvement in South Africa. Mimeo.

Umalusi (2004) Have the standards of the Senior Certificate Examination declined? Summary report on the evaluation of the senior certificate examination. Pretoria: Umalusi.

Yeld, N., Grobler, M., and Sekwane, C. (2004) Investigation into the examination of English Second/Additional Language of selected examination authorities in the years 1992, 1999, 2001 and 2003. Pretoria: Umalusi.

MSCE examination results

Are they a reliable tool for measuring educational standards?

David C Yadidi, Malawi National Examinations Board

Abstract

Over the past few years, there has been a lot of debate in Malawi on the topic of educational standards especially at Malawi School Certificate of Education (MSCE) level.

Rising and falling of MSCE pass rates have both been seen as problems in the public eye. However, the general feeling has been that educational standards have generally declined in the country over the years judging by the low pass rates in the recent MANEB examinations.

Whenever the country has registered a very low pass rate like it happened in 1999, when only 13% of the candidates passed the examination, some quarters of the public have blamed the assessment system used by the Malawi National Examinations Board (MANEB) for being responsible for such dismal examination results. However, when the pass percentage has risen like it did in 2005 with 55% pass rate, again some people have cried foul suspecting MANEB to have "doctored" the results in order to pass many candidates with the intent of pleasing the politicians.

The purpose of this paper is to share with others Malawi's experiences with this controversial topic of standards. Firstly, there will be an attempt to explore answers to three crucial questions relating to standards: what are standards, who sets them and how are they measured? The paper will also describe reforms that the Ministry of Education policy makers have introduced to improve and maintain standards. MANEB's procedures in benchmarking grades will also be scrutinized.

Finally, suggestions will be made on what sort of information on standard setting should be accessed by stakeholders in order to empower them with the necessary knowledge and skills for them to accurately interpret and monitor levels of educational standards other than using examination results alone.

Introduction

The rising and falling of pass rates in the Malawi School Certificate of Education (MSCE) examination have been synonymous with the rising and falling of educational standards in the country.

The topic of education standards has generated a lot of debate in Malawi. Commentators and other interested parties have time and again expressed dissatisfaction with the quality of education in the country.

Because of the low pass rates in the recent years, there have been fears that the Malawian children who are the future workforce may not be receiving the same quality of education as their counterparts in the SADC region and the world at large. If proved correct, this could be detrimental to the socio-economic development of the country. There is also belief by some observers especially those of the older generation that the type of teaching and learning taking place in schools, judged by the dismal MSCE examination results, is inferior to that of the 1960s.

Nyasaland, as Malawi used to be called before independence, used to export secondary school leavers to other countries in East, Central and Southern Africa where they held high positions both in public and private offices. It was believed then that the 'standard' of education in the country was 'superior' to that of her neighbours. Malawi still faces a brain drain but those leaving are with university degrees.

The critics of the current education system have even gone to the extent of blaming the assessment system in the country. They have partly attributed poor education standard to the 'low quality' MSCE examinations administered by Malawi National Examinations Board (MANEB), the country's assessment agency.

The MSCE examination is written after four years of secondary education and is considered as the currency of our education system. Successful students are selected to tertiary institutions or go to other post secondary colleges. Alternatively, they are simply given a school leaver's certificate which they can use to secure a clerical or junior technical job. The MSCE certificate is therefore a passport to success in life.

The problem of the rising and falling of MSCE pass rates

Barely a month after the release of the 2005 MSCE examination results there was a big headline in one of the most widely read local weekly papers. That headline, which appeared on the front page cried, "MANEB SEXED THE 2005 MSCE RESULTS". The allegation was that MANEB, in its bid to win favours from politicians, had deliberately lowered passing marks in order to pass more students than those that actually deserved to pass. The intent according to news paper article, was to make the 2005 results more attractive than past ones, "which had been very embarrassing to government". This allegation triggered protracted debate with some people believing it, while others, especially those familiar with quality of MSCE examinations and with the process of standard fixing mechanism at MANEB, dismissing it as "baseless and malicious". It had to take a "very senior" Government Official

to tell the nation the truth. The official also described the allegation as malicious intended to tarnish the good images of both MANEB and Government. Another allegation followed and MANEB once again denied it in both the print and electronic media. The pass percentage that triggered all this controversy was 55.3.

But why quarrel over 55.3 pass rate?

In the eyes of most Malawians the above pass percentage was too good to be true. According to public opinion, education standard in the country was supposed to be going down, and since according to most people there was no evidence that anything had changed for the better in the schools prior to the 2005 examinations, 55.3 percent pass rate could not be justified.

To appreciate the problem, a comparison of the MSCE pass rates between 1970s and the 1990s to 2004 is being made below:

MSCE Pass Percentage from 1972 to 1976

YEAR	PERCENTAGE OF STUDENTS PASSING MSCE
1972	63.2
1973	54.7
1974	54.7
1975	62.7
1976	65.3

(Source James R, 1976 p28)

MSCE Pass rates from 1993 to 2005

YEAR	PERCENTAGE PASS RATE
1993	46.7
1994	43.1
1995	29.4
1997	30.7
1998	17.9
1999	13.67
2000	19.69
2001	34.86
2002	33.63
2003	41.52
2004	49.70
2005	55.3

(Source: MANEB Statistical Archives)

From the second table it can be noted that academic performance (judged by the pass percentages) has been going down to reach its record low in 1999, with 13.67 percent of the candidates passing. However, there has been gradual increase in pass rates since 2001. MANEB has tried to give reasons for the gradual improvement, citing among other factors efforts of government to improve things and, change of attitude by both teachers and students towards examinations. Cheating which was rampant in the 1990s has decreased in recent years as a result of stringent measures put in place by MANEB since 2001. In 2005

tougher measures were introduced including arresting on the spot those caught cheating. Furthermore, since 2003 examinations have been written in cluster centres i.e. a number of schools writing at one central venue. Examination papers are printed overseas and are no longer kept by individual schools but are kept at police formations and other secure distribution centres. These are collected on daily basis to the examination centres. 'Leakages' of examinations are also a thing of the past because of this arrangement.

It appears therefore that people may not have appreciated the link between these recently introduced measures and pass rates or is due to a general belief that today's learners can not be better than those of the past.

Debate about pass rates elsewhere

Debates about the relationship between pass rates and standards of education are a common feature even in the developed countries. It happened in England when the 2004 A levels registered high pass rates in the top grades and more examinations were passed. Some people became cynical about such good results.

Students, Headteachers and education ministers defended the results. They denied any 'watering down' of the examinations but attributed the positive results to hard work by both students and teachers. Standards minister David Miliband attacked what he called the "national disease" of being cynical about improved examination results (BBC News 19 August 2004 19:23 GMT 20.23 UK).

The Director of the Joint Council for Qualifications Dr Ellie Johnson-Searle confirmed that the results were a true reflection of students' ability.

In Scotland the same year, the reaction to high pass rates was generally positive after the Scottish Qualification Authority (SQA) released the results (Schofied, K in the Scotsman, Tuesday 10th August 2004, Scotsman.com.websites).

However when in 2005 SQA announced equally high pass rates some sectors of the public demanded an independent enquiry to verify the results. The skeptics said a completely independent appraisal of the examination system would ensure "pupils, parents and the business community absolute confidence to the value of qualifications' (The Scotsman, Wed 10 august 2005).

The above examples show that people conclude that the standard of education is lowering when pass rates drop, while on the other hand when there is an improvement in performance they are skeptical about such results.

In order to interpret pass rates and their impact on educational standard more accurately, there is need to fully appreciate the role of assessment in the whole education set up and what impact it has on education. In the Malawi situation, MANEB is often blamed when there are both high and low pass rates. When the MSCE examination registered the lowest pass percentage, in 1999 a Presidential Commission of Enquiry was instituted to investigate

factors that led to such dismal performance. The commission's report dwelt at length on the defects of MANEB as an assessment agency. It cited inadequate financial and human resources, poor processing of examinations and inadequate security. According to the report these defects made it difficult for MANEB to come up with an assessment system which was credible and reliable to assist to enhance learning in the schools. The report also purported that being examination oriented, the education system of the country is greatly influenced by the quality of examinations administered by MANEB. Poor quality assessment policies would therefore result in low education standard.

The problem of defining 'standard'

The term 'standard' has a variety of meanings. 'Standard accommodation' means moderate level of luxury. However, when we say 'holding up a standard' we imply excellence (Cizek, 1991). Others define 'standard' as required expected or accepted behaviour by others in the society. Sakala quotes Costrell (American Economic Review, vol.84 No. 4 of 1993) who defines standards as the required level of proficiency for binary credential, such as high school diploma.

The New Jersey Department of Education defines standard as description of what students should know and be able to do upon completion of a course of study.

Standards provide clear and specific benchmarks for achievement in prescribed areas of study. In America these are developed by panels of teachers, administrators, parents, students and representatives from higher education, business and the community and are revised every five years. However in assessment 'fixing standards' means determining cut scores at various levels of academic achievement.

Standards in the Malawian education context

The history of education in Malawi is associated with the coming of the Christian Missionaries in the early 19th century. However the main objective of the missionary education was to ensure that converts were able to read the Bible and to write and count.

The British colonial government's interest in education only came in 1826 with the establishment of a directorate of education (Chimzimu 1967) following a recommendation of a Phelps-Stokes Commission which urged the colonial Government to take direct control of education and establish schools in the then Nyasaland. The Phelp-Stokes Commission also critised the missionary curriculum, which concentrated on moral education at the expense of other subjects. The Commission developed eight goals of education but these goals mainly centred on the individual and his or her environment. Only primary education was offered then.

It was only in 1940 when the first Secondary school was opened at Blantyre, offering an English Grammar School Curriculum. More secondary schools were opened later on.

When political power was shifted to the Africans in the early 1960s, the approach to and perception of education broadened. Policy makers realized the need to educate large

numbers of young Malawians who would spearhead the political and socio economic development.

Malawi's educational system has therefore undergone transformation from the colonial era when education (then engineered by missionaries) was for evangelism and commercial purposes to the post independence era when it is being seen as an instrument of economic advancement and cultural heritage.

To guarantee quality, education in Malawi is guided by various regulations and instructions. The Ministry of Education which is the responsible ministry initiates all legislation strategies, programmes and projects and provides funds (Ministry of education and UNESCO, 1996).

However there are no specific policy documents that have defined 'standard' in the various curricula. What are there are various official documents outlining government's policy on education. The education system especially at secondary level attempts to equip learners with knowledge, skills, values and attitudes to enable them become responsible and productive adults. The national goals of education which are spelt out in the policy documents cover most of the aspects of life and these have been broken down into seven categories namely;

- Citizen skills,
- Ethical and socio-cultural skills,
- Economic development and environmental management skills,
- Occupational and entrepreneurship skills,
- Practical skills,
- Creative and resourcefulness skills, and
- Scientific and Technological development skills.

(Source: MSCE Syllabi)

However, standards in Malawi are interpreted using pass rates instead of considering all the above aspects. There is also tendency to pack standards against a person's ability to speak and write good English.

Quality control mechanisms

Monitoring of 'standards' in Malawi is spearheaded by the inspectorate division called Educational Methods Advisory (EMAS) Division based at the Ministry of Education head office.

EMAS visits schools to standardize the delivery of education. EMAS teams observe classes and among other details, check schemes of work, lesson plans, and all other lesson aspects including classroom assessments.

In the absence of a National Assessment Board, EMAS is therefore the only reliable monitor of standards in the country. However EMAS reports are mainly for internal consumption by decision makers within the confines of the ministry of education headquarters. The general public is generally unaware of the academic inadequacies uncovered by these inspectors.

The role of the assessment agency in Malawi (MANEB)

Because of the importance attached to certificates the Malawi education system has become examination oriented. MANEB therefore receives a lot of public attention, completely eclipsing other essential agents of the ministry of education such as the Malawi Institute of Education (where the school curricula are developed) and the teacher training colleges. As far as matters of standard are concerned the public eye is focused on examination results.

Of course public examinations are essential. They aim at auditing the curricula, teaching methods and course materials.

Assessments are suppose to provide a feedback on how well students are meeting the expectations of their parents, community and the nation at large. Assessments also provide policy makers with feedback on how well various education policies are working. This feedback is for stimulating changes in education.

MANEB examinations are developed and processed by specially selected specialists following internationally recognized principles of assessment. However the examinations cover mainly the cognitive domains. Although subjects covering affective and psychomotor skills are emphasized in the schools they are minimally tested.

To what extent then does the MSC examination measure educational standards?

It is believed that a good assessment tool will result in good teaching and thereby positively impacting on quality of education. To this effect it is important that assessment agencies like MANEB must ensure that their exams do not distort teaching and learning. They should craft their tests in such a way that such tests do not only focus on issues of technical quality i.e validity and reliability, but also mirror the national goals and objectives. The content matter of particular subjects and the specific objectives of such subjects should also be reflected (Chimwenje, 1997). For assessment to be a reliable tool for enhancing standards, and thereby being a good measure of the same, it should be treated as an integral part of the curriculum, the two being articulated together. Chimwenje (1997) argues that an examination that will successfully measure student's achievement in any subject should relate to the following:

- the nature of the subject and objectives to be achieved,
- the content to be covered,
- the skills and abilities to be fostered, and
- the feeling and learning experiences involved.

In order to ensure credibility MANEB chooses assessment techniques that take into consideration a number of factors including the purpose of the assessment, time and resources available and also the age and competence levels of the MSCE examinees.

Because different subjects syllabi have different assessment demands and description of formats of examination, the assessment by MANEB is varied in nature.

By 1980 when the MSCE examinations were being localized from the Associated Examining Board of the UK (which assisted in the localization process of MANEB), the following were the MSCE types of tests:

1. Aural (Testing Listening comprehension in English and French)
2. Oral (Testing speaking skills—in French only)
3. Theory (written papers in almost all subjects)
4. Practical (in the sciences, Home Sciences, Agriculture and Technical subjects)
5. Coursework/project (in Home and Technical sciences Geography and Agriculture).

However due to high administrative costs the Aural test for English, the course work/projects in Agriculture and Geography had to be phased out, and the number of papers in most subjects reduced.

These limitations therefore mean that MSCE cannot perform some of the skills the examinations were originally intended to test. For example, the absence of the Aural test means students cannot be tested in their listening ability; and the dropping of the projects means no creativity, research skills or original thinking tested.

Furthermore the reduction of papers in the various subjects has narrowed the scope of knowledge to be tested even further, thereby rendering the MSCE examination an insufficient tool for measuring education standards.

The following are some of the problems of using MSCE and other public examination results as monitors of changes in educational standards:

1. When results show a high pass rate it is not certain whether such results are an indication of standards rising or a reduction in the assessment demands. Could it also be that students have become clever? Anything could have happened. The results would not tell us whether if given another test, covering segments of the content or skills not tested in the initial examinations the candidates would score the same grades. Tipia (2000) quoted by Diaz (2000) argues that a student's academic failure is more than his/her failing a test, but more appropriately, it is whether the student's performance is above or below his or her potential. Examination results will therefore not give us much of the above "hidden" information.
2. Quality education should provide new knowledge, skills, values and attitudes to increase people's capacity to participate in the social and economic development of their countries. As the learning process progresses there is need to monitor that learning process to ensure that it is being carried out in accordance with the goals/aims/objectives (standards) laid down by policy makers for positive end results. We know in Malawi EMAS does that but MANEB assessment system is not interested in the learning process itself but rather in the final product. How candidates came by their results and furthermore what will happen to them (candidates) thereafter is not within MANEB's mandate.

3. MSCE examinations are predominately a measure of cognitive achievement with little psychomotor component, which is tested only through very short practical tests. The affective domain is visibly lacking. Such examinations cannot be said to be a true measure of education standards.
4. Passing or failing in the MSCE examination could be a function of many factors and not necessarily as a consequence of the rise or fall in education standards. It could be a problem of the examination testing topics not covered by the majority of the schools. It may also be as a result of the quality of teachers or availability or non availability of teaching and learning materials.
5. Examinations data cannot explain why in a particular year for example, students in one school did better than those of another, despite the schools using the same curriculum, having equally qualified teachers and high calibre students. The disparities in academic achievement may come as a result of the way the examinations were conducted in the two schools. MSCE results simply record the candidate's scores and nothing else. The results cannot therefore be interpreted to mean that the education standard in one school was higher than that in the other.
6. Examinations only sample educational objectives that are measurable. These may not necessarily be more important in education than those that are not measurable.
7. When claims are made that standards have gone up or down compared to past years it is important to justify such claims by explaining the yardstick used. Is the curriculum better now or worse than before? The world is changing socially, economically, politically and technologically, and as such, all new ideas have to be incorporated into a new curriculum. But does that make the new curriculum better? Does it also make the examination that must be changed to be in congruent with the new curriculum more or less difficult? There is need for empirical evidence to arrive at tangible conclusions.
8. Evidence has shown that most schools are in the habit of preparing students on how to pass examinations instead of 'educating' them. Good results may therefore be as a result of drilling. It is also important to take cognizance of the fact that some of the candidates study on their own (externals) and as such their performance cannot be as good as that of school candidates. The more the externals the lower the overall pass rate.
9. As a nation did Malawi define the parameters for identifying an educated person? In the same vein did it define what pass rate would constitute high or low standards? If this is not the case, then 'education standards' as a concept is being used arbitrarily.
10. Monitoring needs to be done early enough so that the findings help the system and the learners. If the MSCE examination were to be useful as a measure of standards it would have to be administered before the completion of the secondary school course. It would then be able to uncover gaps in the knowledge and skills so far acquired by the learners at the time of the examination and corrective measures would be made.
11. Low pass rates may also be a result of tests being flawed if care was not taken at the setting stage. They may be culturally biased against certain categories of the examinees. They may also have tested what the students were not taught. Such examinations would create excessive stress among students.
12. MSCE pass rates went down during the transition period, i.e. between the end of the one party dictatorship and the beginning of the multi-party democracy as a result of increased cases of indiscipline in schools. Democracy might have been constructed as unlimited freedom to students. Many schools were vandalized and therefore kept closed for long

periods of time throughout the year. Some teachers too went on strike at will. The curriculum was intact but instability in the schools affected its delivery thereby resulting in poor pass rates.

13. The introduction of television in Malawi affected the reading culture among students. Pretorious (nd) reports that research has shown that there is relationship between reading ability and academic performance.
14. Examinations are written at the end of a course of study and an exam paper lasts for a short period of time. Examination data do not provide any information on how the students had been progressing to the time of the examination. It is therefore not possible to know whether the results are a true reflection of the candidates' ability or are as a result of external assistance the test takers may have received.
15. Effectiveness of an education system is supposed to be measured by the quality and quantity of its output. Outputs in this respect refer to the school leavers. However, there is no mechanism for predicting how the school leavers are going to satisfy the needs of the society and the nation. Reliance on academic achievement tests as a measure of the value of the outputs is therefore dangerous for it overshadows other factors like their future performance. It has been observed that some people who failed MSCE have proved to be very successful citizens.
16. Studies elsewhere have also shown that students pass or fail examinations as a result of many factors. Diaz (2000) attributes a student's academic success or failure to three elements namely, parents (family causal factor), teachers (academic causal factor) and students (personal causal factor). This therefore means that no matter how well articulated the goals, aims and objectives of an education system may be, if one of the three factors has problems the whole system is doomed to collapse.

Way forward

The following are some of the strategies which policy makers may consider in order to successfully monitor and even enhance education standards:

1. Establish an independent Standards Office. This office would function like OFSTED UK. Members to this body could comprise specialists in assessment, curriculum, industry, information technology and management. The office would be mandated to carry out monitoring of standards in all types of schools in order to identify gaps in the teaching and learning process and suggest corrective measures. Senior managers and teachers of this local 'ofsted' would basically focus on specific targets to raise attainment, while managers from business would provide industrial mentoring to build the students' confidence necessary for a successful future life. Reports of this office would be shared with MANEB which would in turn take into consideration all the points raised in such reports when constructing test items and also when fixing passing scores. Results from such examinations would truly reflect the real academic situation in the schools and would give a true picture of the level of education standards.
2. During standard fixing exercise by MANEB other outside stakeholders should be involved. These could come from industry, Health Sciences and other fields. These would be better placed to advise assessment experts on the expectations of employers from the students. It would also be more helpful if these people could also be involved in setting education standards with education policy makers. With such background

these stakeholders would be in a position to confidently assist in fixing reliable assessment standards.

3. MANEB should diversify its assessment criteria to take into account global concerns about the shortcomings of summative examinations. MANEB could possibly blend its examinations with continuous assessment. If handled professionally and by qualified teachers, continuous assessment would give a better picture of the levels of educational status in the schools than relying on end of the course examinations.
4. MANEB should improve the mode of reporting results to the public to make them more useful. Instead of simply announcing grades and pass/fail information, which does little to help the students, teachers and the community, an attempt should be made to be more detailed. As of now MANEB results do not spell out the candidates' academic strengths and weakness nor do they predict what they will be able to do after passing the examination. It is therefore necessary to make the reporting more transparent.
5. Students preparing to write examinations should be given as much information as possible so that they know what to expect in the examination. Why not? They could even be told in advance which topics questions would come from in the various subjects. They should also be given opportunity to show their best ability in the areas they are strong in, hence the need to include aspects of all the three domains (cognitive, affective and psychomotor) in the examination papers. Test items should be thought provoking, and should arouse the learners' intellect. They should include practical tests that require the students to perform simple experiments, workout puzzles and even solve complex problems.

Conclusion

This presentation has discussed the problem of using public examination results as monitors of change in education using Malawi School Certificates of Education (MSCE) examination as a case study.

It has been argued that examination results alone cannot accurately measure the level of education standard as examination data by themselves cannot provide evidence of whether standards are rising or falling. It has been noted that results may rise due to factors such as the difficult levels of the test, student characteristics, poor teaching and maladministration of the examinations.

It has been recommended that some standards monitoring mechanisms such as the National Assessment Board and the Standards office should be introduced by Government of Malawi. Involvement of various stakeholders in standards setting mechanism has also been suggested.

If the above strategies were adopted, it would be easy not only to monitor but also to improve standards of education in Malawi using MSCE examination.

References

- BBC News (2004) "A levels show rise in top grades" (Thursday 19 August 2004)
- Cizek, G. J. (1999) Setting Passing Scores. University of Toledo)

- Chidalengwa, G. (2004) "A comparative Investigation of students' cognitive motivational characteristics in Conventional and Community Day secondary schools" A dissertation Submitted in partial fulfillment of the requirements of Master of Education (Psychology of Education), University of Malawi.
- Cresswell, M. (2000) "Educational Standards," In *The British academy Review*, January–July 2000 publication.
- Chimzimu, H. D. (1967) "Development of Education in Malawi 1875 – 1967," Final year seminar paper, University of Malawi.
- Chimwenje, C. (1987) "Ensuring the proper articulation of the assessment with the curriculum". A paper presented at a workshop on the assessment of the New Primary School curriculum. Domasi, Zomba, Malawi
- Hansen, ira (No date) "Are the Tests Flawed? Or is Mediocrity the New Standard?" Electric America.
- James, R (1976) Malawi Certificate of Education, Associated Examining Board
- Presidential Commission Enquiry into the MSCE Examination results (2000) Executive Summary of Findings and Recommendations of the inquiry on Malawi School Certificate of Education (MSCE) Examination results
- Pretorias E J (nd) "What they can't read will hurt them": Reading and Academic achievement., www.library.unp.ac.za/innovation/a46 htm.
- Sakala K.M. "Education Standards; A Case of Malawian Schools" a discussion paper presented to MED (Policy, Planning). University of Malawi.
- Schofield K (2005) "Call for Enquiry as exam pass rates rise" The Scotsman Wed 10 August 2005
- Tally and Wang (1999) "Possible causes of and solutions for low Achievement". The Kansas State Board of Education
- The Kansas State Board of education (2000) Possible causes of and solutions For Low Achievement (May 2000)
- Tipia (2000) Personal, family and academic factors affecting low achievement in secondary schools (2000) in Diaz (2000)
- Yadidi D.C. Strategies to Enhance Educational Attainment: A case for MANEB'. A paper intended for the 17th AEAA Conference, Lusaka 1999 (not presented).
- Yadidi D.C. (2004) "Formulating strategies for Removing Imbalances in Academic Performance between Conventional Secondary Schools and Community Day Secondary Schools: The case of the South Eastern Education Division". A dissertation submitted to the University of Derby in partial fulfillment of the Award of Master of Science in Strategic Management degree

Parity of esteem: hope or despair?

Ronel Blom, South African Qualifications Authority

Abstract

The context of an appropriate 21st century education and training system is changing almost as rapidly as modern technology. Questions are being raised, on the one hand, about the ways in which education can become more relevant to the world of work and, on the other hand, how vocationally oriented training could become more valued. With this as the background, the South African National Qualifications Framework promotes the integration of education and training. Quite apart from the historical reasons for the desire to integrate education and training in South Africa, internationally, there seems to be a drive towards a more responsive approach to the needs of society, workplaces and, most importantly, individual learners. One such aspect of this drive is to achieve parity of esteem between different components of the system (i.e. education and training), and the different sites of learning (i.e. institutional and workplace).

This paper will address a number of related issues in terms of parity of esteem between vocational and general education 'standards', including (1) discussing the notion of similarity versus comparability and (2) exploring the notion of epistemology versus relationships. The paper will also explore the extent to which deeply entrenched views of education and training may be inhibiting innovative approaches to systemic, epistemological and curricular changes that could enhance the establishment of comparable, if not equivalent, standards for education and training.

Introduction

The context of an appropriate 21st century education and training system is changing almost as rapidly as modern technology. Questions are being raised, on the one hand, about the ways in which education can become more relevant to the world of work and, on the other hand, how vocationally oriented training could become more valued. With this as the background, the South African National Qualifications Framework promotes the integration of education and training. Quite apart from the historical reasons for the desire to integrate education and training in South Africa, such as the legacy of a system marked 'by division, inequality, segmentation, centralisation and poor accessibility' (Manganyi, 1996:1),

internationally, there seems to be a drive towards a more responsive approach to the needs of society, workplaces, and most importantly, individual learners. This is an on-going debate in South Africa, reflected in the calls for closer links between education and training, parity of esteem between different components of the system (i.e. education and training), and the different sites of learning (i.e. institutional and workplace), greater access to meaningful learning pathways and the improvement of the skills base of the country. Yet, in contrast with the international trend to develop approaches that will unify education and training (or academic and vocational learning) (Raffe, 2005), in South Africa, there seems to be an increasing move away from the first principles of the NQF: of integration, of mobility and meaningful progression through comparable standards, and of enhanced access to education, training and employment opportunities.

This paper will attempt to address these issues from a different perspective. This perspective is one that is derived from questioning the assumptions from which we are proceeding and hopes to explore whether, in fact, we are asking the right questions in relation to integration, standards and parity of esteem.

Four questions will deal with these issues. The first question is—what are our assumptions? The second question will briefly explore current realities. The third asks—what is it that we want to achieve? Finally, the fourth and the most difficult question asks—what are the possible ways to achieve integration, portability, meaningful learning pathways and consequently, parity of esteem?

Parity of esteem

I have read the recently published Umalusi report *Apples and Oranges? A comparison of school and college subjects* (2006), with great interest and an increasing feeling of despair. I found that I had to agree with Young, (2003:10) who says that 'parity of esteem is not a reality in any country'—let alone in a country where the emerging education and training system is struggling to shake off the results of possibly the most pernicious inequalities in education provision in the world. The apartheid system of education and training was not the result of 'benign neglect', but a response to a purposeful and deliberate attempt to keep millions of people 'in their place' (Blom, 2006). Young (1996: 33) points out that parity of esteem 'in theory, guarantee equal opportunities and progression regardless of the learning pathway chosen' and that such a strategy 'would point out how vocational and technical programmes would need to be improved'. In South Africa this takes on a very special meaning, linked to social justice issues, as well as to the quality of education and training. Mehl (2004: 22) says that this stems from the way in which 'society recognizes, rewards and measures learning achievements… It is society that provides ultimate validation of qualifications and accords respect to the bearer. Society awards status and also opportunity and privilege'. The apartheid education and training system skewed these 'customary societal norms' along the lines of race, class and gender.

The findings of the Umalusi study therefore strongly confirm what we know about the weaknesses in our Further Education and Training system—particularly in our vocational education and training system. In Cycle 1 of the NQF Impact Study (SAQA, 2004), for

example, learners at a large public Further Education and Training College took part in a focus group for the Impact Study, and made the comments that:

> Colleges are regarded as inferior.
>
> [and]
>
> We have a problem with doing N3 courses, because when we approach a university, they tell us that the N3 is not a proper Matric and that we cannot be admitted.

Likewise, lecturers at the same college noted:

> The NATED programmes are not relevant. We are not offering programmes that are relevant to learners. Maybe industry is not pushing us hard enough.

The Umalusi report confirms these views and notes that in fact, the technical college qualifications (vocational qualifications) were originally 'designed to be an easier alternative' (Umalusi, 2006: 1) and were therefore never meant to have the same esteem as a general/academic Higher Grade matriculation.

The extent to which the apartheid education and training system has impacted on the quality of learning, the status and esteem of graduates, is evident to this day, 12 years after the first free elections in 1994. At the Portfolio Committee meeting of Parliament on the 15th of May 2006, for example, in light of discussions about the critical skills needs of our country a member of the Parliamentary Committee, asked the Minister of Education, Naledi Pandor:

> …what [is] being done to correct the perception that those who choose vocational training are of lesser intelligence…?

This brings us to the sub-theme of this conference: *thinking about standards over time, across subjects, and across learning paths*. The Umalusi report suggests that standards between general/academic schooling and vocational college education are not comparable. These findings lead to a number of recommendations important to a quality assurance body, but I am of the opinion that the most important findings of the study are not made explicit.

In my view the most important finding is: **at present standards are not comparable; currently, parity of esteem is not a reality**.

However, what struck me is that the Umalusi report is perhaps a vindication of deeply entrenched views of education and training which seem to inhibit the development of innovative approaches to systemic, epistemological and curricular reforms that could enhance and support the establishment of comparable, if not equivalent, standards for education and training. The report seems to suggest that the NQF is at fault, that the NQF has not had an impact on vocational education in South Africa. Of course, one can argue that this is entirely correct—but not for the reasons that the report puts forward.

Van Rooyen, (2006:2) notes, for example, that

> …over the past 25 years vocational education has been suffering from neglect and gross mismanagement by both the present and past governments. Although the technical colleges were administered provincially, changes in curriculum were usually initiated from the national government offices. The curriculum for technical occupations were never reviewed [for more than] 20 years and institutional (practical) training was totally neglected.

And this brings me to the questions I want to raise in this paper:

- What seem to be the assumptions?
- What are the realities?
- What do we want to achieve with our new education and training system?
- What are the ways in which we can achieve these?

What are the assumptions?

The most important assumption is that any teaching and learning in South Africa in the further education and training (FET) band, i.e. in schools and vocational colleges, at present adequately prepare learners for further and higher education[1].

This assumption should be questioned. The Third International Mathematics and Science Study (TIMMS 1994/1995) and its Repeat (TIMMS-R 1998/1999) for example, found from a descriptive analysis of 194 South African schools and 8141 pupils that (Howie, 2002: 4):

> South African pupils performed significantly worse [in mathematics and science] than all the other participating countries in TIMMS-R including other developing countries.
>
> There is no significant difference in the performance of the South African pupils in 1998 [the Repeat] and those in 1995 [the year of the first assessment].

While it must be acknowledged that much has certainly been done in schools in the 8 years following these results, most notably in the areas of infrastructure and increased enrolments (Jansen, 2002), it should be noted that the new National Senior Certificate (NSC), (the FETC General/Academic) is only to be implemented for the first time in 2008. In addition, the 'old' Senior Certificate examination has been in steady decline since 1997 (SAUVCA, 2001), meaning that less learners reach 'Matric', despite improved enrolment figures at the lower end of the system. In addition, in a report commissioned by SAUVCA (now HESA) on *The Challenges of Access and Admissions* (2001: 13) it is noted that

> The pressure on [provincial] departments [of education] to improve their overall Senior Certificate pass rates encourages them preferentially to register students for non-matriculation endorsement subject combinations, which has translated into a smaller number of higher grade candidates, with concomitant increase in standard grade. This has led to an overall increase in the higher grade pass rate. Alongside this, the performance in standard grade is very poor.

[1] The Umalusi report explicitly does not attempt to address the occupationally based qualifications and these are therefore excluded for the moment.

Consequently, the HE sector feels constrained to increase participation in university study by the 'large number of learners who are underprepared' and 'the lack of reliability or trustworthiness of matric results' (SAUVCA, 2001: 20, 21).

It is therefore simply still too soon to say whether the new (NQF) qualification will improve learners' preparedness. Not even the Higher Education (HE) sector, as the recipients of learners graduating with the new NSC, can predict whether this will be the case. Instead, HE has consistently noted that the current matriculation certificate is not a good predictor of success in higher education.

However, the same cannot be said of the FET colleges. The new Further Education and Training Certificate (Vocational), which is meant to replace the old NATED qualifications offered at the colleges, is in its final stages of development as we speak. The Umalusi study therefore compared vocational qualifications and curricula (and assessment regimes), which are widely acknowledged as out-dated, weaker even than current school subjects, and irrelevant to the world of work, with current matriculation subjects—which are also subject to change. It is therefore fair to say that:

> …the lack of progress in developing the new college-oriented FET Certificate makes it impossible to judge the depth and power of the Department's educational vision for the [vocational] sector (McGrath, 2002).

In addition, many studies have suggested that colleges have considerable room for improvement. A National Business Initiative (NBI) survey indicated that 'the overall pass rate [in colleges] nationally stood at 53 per cent in 2000 and throughput rate at 47 per cent' (Powel & Hall, 2002: 77, in Mc Grath, *et al*).

Further, the language of instruction for 79 per cent of the student population in 2000 at the time of the NBI study would have been a second or third language. The TIMMS-R results for example, indicate that 'language of learning was found to be a significant predictor of pupils' achievement in South Africa; and those classes where the language predominantly used was the same as the language of the test, achieved higher mathematics scores' (Howie, 2002: 5). This is not the case for the majority of current learners in the vocational education sector.

I am therefore of the opinion that the weaknesses in current vocational education provisioning has very little to do with the South African NQF—it has more to do with the fact that we allowed the perpetuation of an old system which did not meet the needs of any of the three categories of learners in the vocational education market: the pre-employed; the currently employed; and the unemployed (Kraak, 2004):

> …technical colleges provided poor pre-employment training with very low placement rates. In addition, the apprenticeship system, as a form of employment induction, is in severe decline (p. 120).

Yet, the Umalusi report states that (2006: 2):

> …it was hoped that the status of vocational education would be raised through the introduction of an ambitiously designed National Qualifications Framework (NQF). By designating that qualifications were all at level four on the NQF, the level of Grade 12 qualifications, it was thought that the status of all qualifications at that level would be the same and that all qualifications at level four could lead to higher education programmes…

My strong contention is that it is impossible to say whether the status of vocational education will be raised, simply because the system has not yet implemented the new qualifications, curricula, learning programmes and articulation routes needed to enhance parity of esteem.

Minister Pandor, for example, stated in the debate on the 15th May Parliamentary Portfolio Committee that

> Further work needed to be put into educating the public, as the perception of FET Colleges as "lower-class" was indeed disappointing, and to ensuring that the teaching at higher education institutions was more closely aligned to allow greater mobility.

What are the realities?

The reality is that the skills crisis has arrived; that the ideal of 'vital intermediate to higher-level skills and competencies the country needs to chart its own course in the global competitive world of the 21st century' (RSA, 1998:14) has not been achieved—we have been 'sleeping while the beds are burning' (Talking Heads). Minister Pandor, for example, reported at the Parliamentary Portfolio Committee that the FET colleges Bill 'was [only!] gazetted last week as part of the plan to re-cast colleges on the critical edges of high skills training'. Yet, Chisholm already noted in 1992 that

> …the South African system of vocational training provision has always been characterised by a weak and fragmented education-led, college-based system and an almost non-existent employer-led work-based system. Added to this, issues of economic growth and the development of high level skills have always been absent features in the story of technical and industrial education provision in South Africa (in Badroodien, 2004: 44).

The opportunity to reform this situation and the promise that vocational education in South Africa held for the economic and social development of the country and for individuals in the late 1980's, was constrained by the political situation of the time, a fragmented qualification structure for the sector and 'low trust' between the social partners (Kraak, 2004: 47). Coupled with the lack of policy coherence and co-ordination of training efforts, the vocational sector was in need of critical reform. In 2006 this is still the case: the articulation of education and training courses is still constrained; there is still 'an absence of links between the training system and the formal education system'. In fact, if the report *An interdependent National Qualifications Framework System Consultative Document* (DoE & DoL, 2003: 14) is anything to go by, then the links still have to be established:

> Learning pathways cannot be sealed off from one another, as though a learner is fated to stay on one route once a choice has been made. The principle of flexibility must ensure that links are available for learners to move from one pathway to

> another, to be credited appropriately with learning achievements that are relevant to the new pathway and be afforded the opportunity to acquire additional learning that would enable the learner to make an efficient transition

Learners in the public FET college focus group for Cycle 1 of the NQF Impact Study (SAQA, 2004) noted 'we don't need these courses anyway, as people without these courses can get the same jobs' and '[the colleges] do not give us all the required training to work in large companies'.

I am of the opinion, and the Umalusi report confirms this opinion, that the vocational education sector is still in trouble.

But what are some of the other realities, particularly in the so-called occupationally based sector? Hundreds of occupationally directed qualifications have been developed, including a large number of Further Education and Training Certificates (FETCs). Hundreds of learnerships have been developed and implemented. Yet, the status of these achievements is uncertain, particularly from Umalusi's quality assurance point of view. But so too is the status of the vocational education system. Kraak (2004:121) notes that

> …the previous apprenticeship system, which had very loose requirements regarding the linkage between theoretical training and work experience. Most often in the past, apprentices would undergo a minimal level of theoretical training at a technical college (acquiring certificates N1 to N3 [equivalent to Grades 10 to 12], which were often unrelated to their practical training) with little supervision or structured induction into skilled work at their places of employment.

In a pilot study undertaken by the National Access Consortium Western Cape (NACWC) (2001: 87) the weaknesses of the current FET sector were confirmed yet again:

> …in the college sector, sharp differences can be discerned between 'practical' subjects and 'theory' subjects in terms of methodology, classroom practice and assessment. Theory-based courses tend to be characterised by and limited to delivery from teacher to student, and are dominated by national examinations that determine who passes and who fails. [Students undertake informal hours in a workplace and] the college therefore plays no role in monitoring the student to see who eventually qualifies or not, or in seeing how their delivery has impacted on the learner's workplace performance.

But where, in the pilot, Engineering courses, which 'traditionally effectively separated theory and practical training, were brought together into a year-long programme that integrates theory and practice', and combined this programme with communications, computer literacy and entrepreneurship, in a precursor to the learnership system, it became evident that the range of skills learners would gain from the qualification was considerably broadened (NACWC, 2001).

However, public FET colleges are not encouraged to offer occupationally based qualifications, and certainly will not be funded by the Department of Education to offer such. This means that the colleges, which are well placed to develop a responsive sector to the world of work, will be constrained in their contribution to the emerging occupationally

based routes to qualifications—resulting possibly in yet again, stratified statuses for education and training, with the National Senior Certificate at the top, the FETC Vocational in second place, and the occupationally-based FETCs at the bottom of the rung.

Is this what we want to achieve?

I am of the opinion that the FET sector finds itself in the unique space where it can prevent new systemic disparities between the sub-sectors. In other words, the sector is in flux, major changes are on their way. This space offers important opportunities to lift the value of particularly, vocational and occupationally based education and training. Now is not the time to systemically entrench an old system by disparaging some components of the system and by producing a report, which smacks of a particular bias, based on an acknowledged outdated system in need of crucial reform.

If we could therefore firstly acknowledge that there is much work to be done, in all of the sub-sectors of further education and training, then we can start developing innovative ways of complementarity. According to the Consultative Document (DoE & DoL, 2003: 6, 7):

> Educators recognise the importance of career preparation but stress that preparation for work must be embedded in programmes of much broader educational value for individual and social development…Workplace learning practitioners elevate the importance of work for sustaining and enhancing life and society, recognise the value of fundamental education, but argue that work-readiness and work competence are best acquired through learning that is embedded in real work experience.
>
> …the tension between these positions has seemed at times to polarise viewpoints and exasperate relations between the protagonists. But each perspective…has merit. They are not in fact opposites, but equally essential facets of the same national learning system. The National Qualifications Framework is a vital mechanism for holding the tension between them and bringing out the complementary and mutually reinforcing attributes of institutional and workplace learning. It is necessary to stand above simplistic dichotomies and attempt to understand more holistically the interface between the two worlds especially since both are undergoing rapid change…The further development of the NQF can be approached in such a manner that respects the different modes of learning and encourages collaboration and inter-dependence…without compromising the unique value each learning perspective brings to the whole.

What is it that we want to achieve?

The value of the Umalusi report, in my opinion, is that it creates an awareness of a crippling symptom of the education and training sector of the apartheid era. However, it is clear that the problems facing the Further Education and Training (FET) band are much more wide-ranging and potentially may have a much greater impact on our emerging system. An evaluation of the FET band needs to take into consideration all the contextual factors that may influence our decisions, including an acknowledgement that much work needs to be done, and that nothing will happen by itself. The people and organisations who have been given the responsibility for our emerging system, will have to approach the problems of the

FET band from a different perspective, including looking at factors such as learning, delivery, epistemology, articulation, progression and parity of esteem in relation to the needs of individuals and society—and then seek ways to align the whole sector with these needs.

The Departments of Education and Labour, in their initial joint response to the *Report of the Study Team on the Implementation of the National Qualifications Framework* (2002: 15), say that 'a Further Education and Training Certificate (FETC) must equip learners for further learning either at work or in higher education' and propose three pathways: General, General Vocational, and Trade, Occupational and Professional (TOP) (2003: 15).

The theory is that the General 'track' and the General Vocational 'track' enables 'portability of learning credits…to permit articulation with other learning pathways' but the report cautions that '[i]t has not yet been clarified how the FET schools programme will articulate with that of the colleges' (2003: 15). Likewise the TOP pathway 'would in principle permit workers to achieve a level 4 qualification and proceed beyond that current glass ceiling to level 5 and subsequent qualifications without leaving the workplace' (2003: 16). To ensure that the pathways are 'not walled off from the next, an articulation column is created between them to enable vertical, horizontal and diagonal articulation between qualifications' (2003: 17).

According to the Departments the FET system will therefore serve the pre-employed, the employed and the unemployed 'who are seeking to enter or progress in or change a career pathway, or equip themselves for admission to higher education, or both' (2003: 14).

Diagrammatically, this is represented as follows (DoE & DoL, 2003: 17):

Table 1: Proposed revised National Qualifications Framework

NQF Band	General/academic	Articulation	General/vocational	Articulation	Trade/occupational/ professional
HE	Discipline-based	Credits	Career-focused	Credits	Occupational recognition or context-based workplace qualifications
FE	Discipline-based	Credits	Vocational qualifications	Credits	
GE	General education qualifications				

How is articulation, progression and parity of esteem and consequently, the recognition of credits attained in different parts of the system, to take place? What credits could be transferred between the 'tracks'?

Table 2: Articulation and credit transfer routes from general/academic to occupationally-based qualifications

NQF Band	General/academic	Articulation	General/vocational	Articulation	Trade/occupational/ professional
HE	Discipline-based	Credits	Career-focused	Credits	Occupational recognition or context-based workplace qualifications
FE	Discipline-based	Credits	Vocational qualifications	Credits	
GE	General education qualifications				

Theoretically then, a learner who started off in the general/academic track, but who wanted to move towards vocational qualifications, could move horizontally at the same level and/or diagonally into career-focused HE. Likewise, a learner who finds him/herself in the vocational track, could move horizontally at the same level and/or diagonally into a context-based HE qualification. This is quite straightforward and I doubt that anyone will dispute mobility of this nature. However, will it also be possible for a learner to move from the opposite end 'of the same national learning system' (DoE & DoL, 2003: 7)?

In other words:

Table 3: Articulation and credit transfer routes from to occupationally-based qualifications to general/academic qualifications

NQF Band	General/academic	Articulation	General/vocational	Articulation	Trade/occupational/ professional
HE	Discipline-based	Credits	Career-focused	Credits	Occupational recognition or context-based workplace qualifications
FE	Discipline-based	Credits	Vocational qualifications	Credits	
GE	General education qualifications				

If this is not what the Consultative Document means with the statement that the new system will enable all learners 'who are seeking to enter or progress in or change a career pathway, or equip themselves for admission to higher education, or both' (2003: 14), then I want to suggest that we are not serious about articulation, meaningful progression and parity of esteem through comparable standards, and of enhanced access to education, training and employment opportunities for the full spectrum of learners in the system.

Much education and training provision already reflect that education and training are not opposites. The Consultative Document (DoE & DoL, 2003: 21) reflects this in the discussion of the three pathways and the qualifications within them:

- Partnered pairs of qualifications, for example 'occupational and professional training in many fields exemplifies this model'
- Stand-alone discipline-based qualification with a component of workplace practice, for example the 'co-operative education model'
- Stand-alone occupational context-based qualifications with a component of discipline-based study, for example 'learnership programmes illustrate this type of partnership'.

Diagrammatically, this is represented as follows:

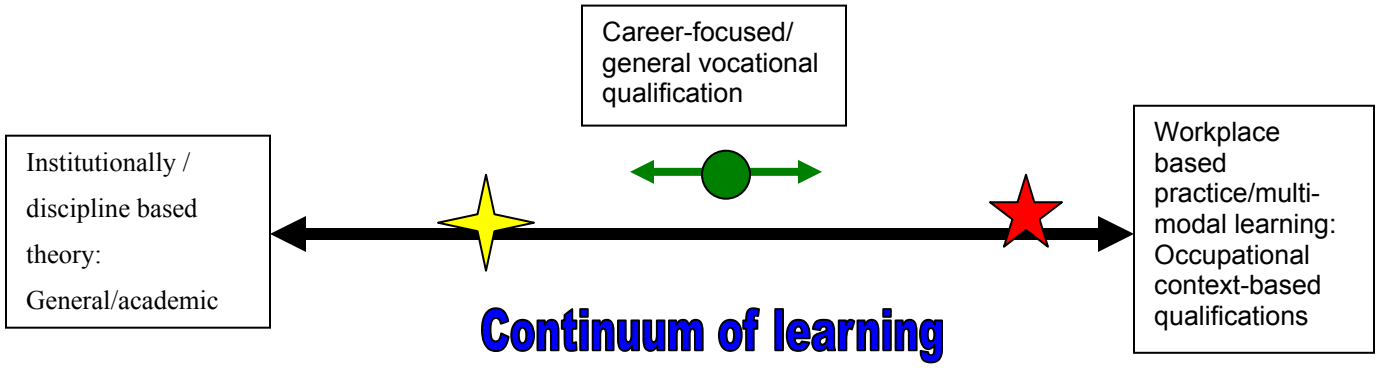


Figure 1: **A continuum of learning**

The extent to which the purposes and rationale of a qualification are defined by the pursuit of discipline based learning (education) or by the utility value in the workplace (training), places a qualification (or set of related qualifications) in a particular place along the continuum. The yellow star in Figure 1, for example, indicates that a particular qualification is mostly about the development of discipline-based knowledge, but with some tentative links to the world of work. The red star, for a qualification at the opposite end of the continuum would therefore be much more occupationally oriented.

However, there are not only these two extremes. A third dimension: a 'career-focused' or 'general vocational' qualification, which 'looks both ways' is also conceptualized (DoE & DoL, 2003). This type of qualification is reflected as the green arrow (Blom, 2006).

I believe it is time to start asking the harder questions:

Are the difficulties we are experiencing in achieving parity of esteem as a result of 'a positional good' (or put differently - perhaps power struggles and vested interests)?

A positional good is one whose value declines as other people have more and more of it. In a credentialist society qualifications are a positional good, and a reform which aims to raise the status of vocational [and occupationally-based] qualifications (and increase their value) threatens to undermine the positional value of academic qualifications. It therefore

threatens the institutions which deliver academic qualifications and the social groups which most often achieve them (Raffe, 2005: 27).

Dare I ask whether the apparent unwillingness to look at alternatives to the traditional organisation of education and training is perhaps simply power play? If this is not the case, or even if it is, should we not seriously consider that it is 'necessary to stand above simple dichotomies and attempt to understand more holistically the interface between the two worlds' (DoE & DoL, 2003: 7)?

How do we achieve parity of esteem?

Perhaps it is time to acknowledge that we need to look at these issues with a different lens—to acknowledge that we cannot hold onto views that fly in the face of international trends, that if we want to solve the skills crisis of the country, and of unemployment and lack of progression, we cannot afford for vocational and occupationally-based qualifications to be considered for those people of 'lesser intelligence'. We have to find ways of raising the 'positional good' of all qualifications. This will not be possible without deliberate, sustained effort from all involved.

This effort will have to take place at all levels of education and training. On a macro level, we need to do the long-overdue analysis of legislation governing the different components of education and training and undertake the clarifications and interpretations that will take the system forward, as opposed to being used as a tool to say 'who is the boss?'

On a meso level, we need to ensure that the qualifications themselves contain internal consistencies and specify transferable credits. Theoretically, all Further Education and Training Certificates (FETCs), namely the new National Senior Certificate: Schooling; the National Certificate: Vocational and the FETC: Occupationally-based, have a comparable structure: Fundamentals, Core and Electives, with the Fundamentals being the base and minimum benchmark. We therefore need to collectively build in, and specify credit transfer and articulation possibilities.

On a micro level, we have to make sense of learning outcomes, curricula, learning programmes and assessment regimes in order to make meaningful comparisons. Crucially, we need to capacitate education and training practitioners.

Then we have to strengthen, not disparage, new areas of knowledge-production. Experts must get involved in seeking to strengthen poorly defined areas of knowledge production and establish a functional and knowledge taxonomy in sectors where strong disciplinary based traditions do not exist (Mehl, 2004).

On all levels, we have to build trust, we have to share practice, we have to think in terms of the system and the learners within them. We have to build quality assurance mechanisms that will enhance learning—wherever this may occur—and recognise the value and contribution it makes to the economic and social development of the country as a whole.

We have to ask ourselves the most difficult question: Have we done enough of the hard thinking to do the hard work? I do not believe we have.

Conclusion

Albert Einstein said:

> You cannot solve a problem with the same thinking that created it.

In the light of this famous statement, perhaps parity of esteem needs to be clarified. In my view, it does not mean that the widely differing purposes of qualifications should be ignored or devalued. It also does not mean that curricula, learning programmes, assessment regimes, modes of delivery will be the same, or even similar. It does, however, mean that as a system, we value all learning, regardless of where it was attained and that the system therefore, should not entrench a previously (discredited) view of education, training and workplace learning. Instead, the system should enable meaningful comparisons of standards over time, across subjects, and across learning paths.

I want to argue that the problem is with the model, reflecting a particular bias, not the principle of parity of esteem. Raffe, (2005: 26) says

> The argument is further confused by being polarised in terms of education and training—or rather, in terms of *caricatures* of education and training.

Is parity of esteem an intractable ideal? Perhaps, but I personally will not believe that this is the case until all avenues have been explored.

So, the first assumption dealing with the FET band should be that there is much work to be done—including the work that will enhance parity of esteem, but even more importantly, the work to raise the quality of all FET education and training.

This assumption is based on the realities of the sector that has somehow remained neglected—9 years after the introduction of a National Qualifications Framework. Our work should therefore be focused on what it is that we want to achieve—and how this may impact on individuals' lives, the communities within which they find themselves, the economy that sustains such communities, and the wider social justice goals of transformation, redress and improved access to education and training and employment opportunities.

Perhaps, at this point, it is time to remind ourselves of the goals of our emerging education and training system and remember why these goals were important at the time. If the goals of an integrated, high quality education and training system that will facilitate access, mobility and progression for the individuals in the system in order to achieve their full personal development, no longer holds, then we should develop new objectives. If, however, these are still true, then we have to find ways in which to make this possible.

I agree with Raffe (2005) that it at this juncture, it is time to 'identify and overcome the barriers to progress, to plan the evolution of the qualifications framework, and of the wider

education and training system…and to manage and steer this evolution. In other words, the…challenge is to envision, motivate and manage the change' as partners, not opponents, in South Africa's education and training system.

References

- Blom, R. (2006) The South African National Qualifications Framework – an integrative and socially-cohesive approach towards valuing qualifications and skills. Seminar paper for the Scottish Executive Education Department, 25 April 2006.
- Blom, R.. (2006) The Concept(s) of an Integrated Qualifications Framework: A Conceptual Framework of the uses and meanings of Integration in the South African Education and Training System. Seminar paper for the University of Edinburgh, 25 April 2006.
- Chisholm, L. (1992). South African Technical Colleges: Policy options. Johannesburg: University of the Witwatersrand, Education Policy Unit.
- Department of Education and Department of Labour (2002). Report of the Study Team on the Implementation of the National Qualifications Framework. Pretoria.
- Department of Education and Department of Labour (2003). An interdependent National Qualifications Framework system. Consultative Document. Pretoria.
- Education Portfolio Committee (2006) Accessed 9 May 2006 from http://pmg.org.za/docs/2006/060509pandor.ppt
- Howie, S. (2002) English Language Proficiency and Contextual Factors Influencing Mathematics Achievement of Secondary School Pupils in South Africa. Summary. Thesis University of Twente, Enschede.
- Jansen, J.D. (2002) The Sustainability of Education Reform in South Africa: A Critical Assessment, 1994–2002. Report prepared for the United Nations Development Programme.
- Kraak, A. (2004) The National Skills Development Strategy: a new institutional regime for skills formation in post-apartheid South Africa. In McGrath, S., Badroodien, A., Kraak, A. and Unwin, L. (eds). *Shifting Understandings of Skills in South Africa: Overcoming the historical imprint of a low skills regime*. HSRC Press. Cape Town.
- Manganyi, N.C. (1996) The South African Qualifications Authority Act. In the Inter-Ministerial Working Group Conference on the National Qualifications Framework. Proceedings. Johannesburg.
- McGrath, S., Badroodien, A., Kraak, A. and Unwin, L. (eds). (2004) Shifting Understandings of Skills in South Africa: Overcoming the historical imprint of a low skills regime. HSRC Press. Cape Town.
- Mehl, M. 2004. 'The National Qualifications Framework: Quo Vadis?' In *SAQA Bulletin*, 5 (1) pp. 21–46.
- National Access Consortium Western Cape (2001) Opening the doors of Learning. From Policy to Implementation. Cape Town.
- Powell, L. and Hall, G. (2002) Quantitative overview of the Further Education and Training College sector: The new landscape. Pretoria. Department of Education.
- Raffe, D. (2005) 'National Qualifications Frameworks as integrated qualifications frameworks'. In *SAQA Bulletin*, 8(1), pp. 24–31.
- South African University Vice Chancellors' Association (2001) The Challenges of Access and Admissions. SAUVCA, Pretoria.
- Umalusi (2006) *Apples and Oranges? A comparison of school and college subjects*. Council for Quality Assurance in General and Further Education and Training. Pretoria.

- Van Rooyen, M. (2006) Comparing apples and oranges: Comment on the Umalusi research. Cutting Edge. May 2006.
- Young, M. (1996) Theoretical grounding and an international perspective. In the Inter-Ministerial Working Group Conference on the National Qualifications Framework. Proceedings. Johannesburg.
- Young, M. (2003) *An Interdependent Qualifications Framework System*. A Consultative Document prepared by a Joint Task Team of the Departments of Education and Labour. Report prepared for the Council on Higher Education.

Can we ensure that vocational programmes are at the same 'standard' as general education programmes?

The case of Geometrical and Orthographic Drawing in Malawi Craft

Gerald Axel Chiunda, Malawi National Examinations Board

Abstract

Malawi Craft is a vocational programme that was established to train school leavers in ten technical trades. School leavers who join craft courses may or may not have a background in Geometrical and Orthographic Drawing (G & O). Nevertheless, G & O appears to be a basic component of all the trades under Malawi Craft.

The aim of this paper is two-fold: first, to compare the content of G & O at Senior Secondary School level with that at Malawi Craft level which is a post-secondary. In case of G & O being more at the secondary school level, suggestions will be made for improvement. The first part will end with conducting statistical analysis of comparative performance in college between those who did G & O in school against those who did not.

Secondly, the paper will analyse cut-scores for various grade levels in order to establish whether or not assessment standards have been maintained over the years in subjects where G & O is a component.

Introduction

Malawi Craft is a vocational programme that was established to train school leavers in ten trades namely:

- Motor Vehicle Mechanics,
- Machine Woodworking,
- Mechanical Trades (General Fitting),

- Brick work,
- Carpentry and Joinery,
- Plumbing,
- Welding and Fabrication,
- Vehicle Body Repair,
- Painting and Decorating, and
- Electrical Trades.

In all these trades, students train at craft (elementary) and Advanced Craft levels. Up until 2002, The Malawi Polytechnic used to train students at technician level, a step between Advanced Craft and a Degree level of craft.

School leavers who join craft courses either have a background in Geometrical and Orthographic Drawing (G & O) commonly known as Technical Drawing (TD) or they do not. Nevertheless, TD appears to be a basic component of all the trades.

Research questions

In order to appreciate the bigger question, the title of this paper, TD was used as decider variable for the following specific questions:

1. How much of TD is there at MSCE level?
2. How much of TD is there at Craft level?
3. Do students who did TD at school level perform better in colleges than those who did not?
4. Have grading standards in subjects that have TD component been maintained over the years?
5. How much of TD should be at school and college levels?

Methodology

In order to answer the research questions, desk research and subject matter specialists were used.

The desk research involved analyzing syllabus content of TD at both school and college levels, to which two subject matter specialists established the depth of the TD. Furthermore, multiple regression was done to establish the extent to which grading standards had been maintained from 2002 to 2004 Craft Examinations.

Results

Content Coverage of TD at MSCE level

At MSCE level, TD is examined as a stand-alone subject and it comes in two papers, Paper I and II. A summation of the content in the two papers is the entire syllabus at MSCE level. There also a coursework which constitute 20% of the overall grade. The syllabus aims at:

1. developing student's ability to understand and appreciate good draughtsmanship;
2. developing students' skills in the use of drawing equipment; and

3. developing students' ability to use geometrical constructions and orthographical drawing in producing solutions to practical problems. The syllabus is covered in two years, and specially trains students in basic geometric and orthographic drawing skills in line/letters, basic geometric construction, scales, triangles, circles and parts, quadrilaterals, polygons, tangents to a circle, enlargements, areas, normals, loci, cycloid, helix, parabola, hyperbola, blending radii, construction of an ellipse, pictorial drawing, and orthographic projection.

Content coverage of TD at craft level

The aim of the TD syllabus at craft level is for communication; that is students should be able to read and interpret a working drawing.

As discussed above, TD is a component in almost all the trades at craft level at which candidates sit two papers whose scores are combined with those of practical component. At either craft or advanced craft, there are therefore three components namely:

- Technology,
- Associated paper and
- Practical commonly known as phase tests.

Technology

This component involves sketching and little calculations. The skills assessed in this component include use of materials, site work, way of carrying out certain tasks, etc. In a nutshell, the technology is a paper that assesses the theoretical grounding of a candidate in a particular trade.

Practical

In this component, candidates work on a project. The candidate is assessed on using his/her hands to apply learnt theory.

Associated paper

This paper assesses candidate's skills in Science and TD. The coverage of TD mainly focuses on the applied drawing detail of each trade. The depth varies from trade to trade. Tables 3.2.3 (a-h) has details about TD coverage at advanced level by year and by question for some trades.

Table 1 Carpentry and joinery

Year	Topic	Question number	Marks
2002	Development of surfaces	9	5
2003	Development of surfaces	9	5
2003	Development and showing of details	10	5
2004	Orthographic drawing	9	5
2004	Development and showing of details	10	5

Table 2 Brickwork

Year	Topic	Question number	Marks
2002	Orthographic and Geometrical Drawing	7	5

	hyperbola		
2002-2004	True length, plan of frustum, development of prism	8	5
2002-2004	Setting out of arch, drawing an arch	9	5
2002-2004	Isometric and oblique projection, elliptical arches	10	5

Table 3 Motor vehicle mechanic

Year	Topic	Question number	Marks
2002	Isometric views	10	10
2003	Isometric views	10	10
2004	Isometric views	10	10

Table 4 Machine woodworking

Year	Topic	Question number	Marks
2002	Geometrical	19	4
2002	Freehand drawing	20	4
2002	Orthographic	21	4
2002	Orthographic	22	4
2002	Orthographic	23	4
2002	Development	24	4
2002	Pictorial	25	4
2003	No examinations		
2004	Orthographic	20	8
2004	Development	21	8
2004	Pictorial	22	8

Table 5 Mechanical trade

Year	Topic	Question number	Marks
2002	Orthographical, projection, assembly drawing, Ditto	15	20
2002	Orthographical, projection, assembly drawing, Ditto	16	10
2003	Orthographic drawing, assembly drawing, Ditto	15	10
2003	Orthographic drawing, assembly drawing, Ditto	16	20
2004	Orthographic drawing, assembly drawing, Ditto	15	20
2004	Orthographic drawing, assembly drawing, Ditto	15	20

Table 6 Plumbing

Year	Topic	Question number	Marks
2002, 03, 04	1ST angle	12	5
2002, 03, 04	Surface development	13	5
2002, 03, 04	Hexagon	14	5
2002, 03, 04	Ellipse	15	5

Table 7 Welding and fabrication

Year	Topic	Question number	Marks
2002, 03, 04	1ST angle	1	10
2002, 03, 04	Orthographic	2	10
2002, 03, 04	Development	3	10
2002, 03, 04	Circles, tangents, sectors	4	10

Table 8 Vehicle body repair

Year	Topic	Question number	Marks
2002	Instruments for drawing	5	10
2002	Orthographic, 3rd Angle	1 2 3	10 10 10
2003	Isometric view	11	10
2003	Development	12	10
2004	Constructing a Δ	1	3
2004	1st Angle orthographic	2	10
2004	Surface development	3	10
2004	Circle (tangent, chord, segment)	4	3

It would appear that coverage of TD mainly focuses on applied drawing detail of each trade. The depth varies from one trade to another, and is much more detailed than at MSCE. The coverage at MSCE is basic, generalized and spread over many areas.

Looking at Craft within the period under study, item writers and moderators maintained the total number of marks allocated to drawing questions in Carpentry & Joinery, Brickwork, Plumbing and Motor Vehicle Mechanics. However, there were seemingly variations in marks allocation in Machine Woodworking and Mechanical Trade. As for Machine Woodworking, its table of Test Specification shows that the total marks for drawing questions is 20% of the whole Associated paper. The variation in the allocated marks in 2002 and 2004 was +8 and +4 respectively. This shows that the setters and moderators did not follow the table of test Specifications. The same picture could be true of Vehicle Body Repair. There was variation in the marks allocated in 2002, 2003 and 2004 of +2, -5 and -2 respectively.

Statistical comparison in performance between those who did G & O in schools and those who did not

Scores of the first three college tests from three Technical colleges were analysed using T-test, linear regression, and one-way ANOVA. Out of the 50 students in the three colleges, only 5 did G & O in school. Results of T-test for Test 1 between those who did G & O and those who did not in school indicated a mean difference of 1.90 which was significant (i.e. p-value = 0.000) at beta 0.05. The small mean difference was supported by R-value of -0.213 using linear regression. Pearson correlation between scores of the first test in college and those who did G & O gave a p-value of 0.146. That is, the correlation was not significant. Finally, One-way ANOVA between G & O and the three tests showed p-values of 0.146, 0.054 and 0.404 respectively for test one, test two and test three.

Maintenance of assessment standards over the years

In order to establish whether or not assessment standards have been maintained over the years, analysis should ideally be focused on composition and structure of the test papers, and cut-scores over a period of time. Detailed analysis of the past papers was not possible because of technical limitations. However, analysis of cut-scores for associated paper of Brickwork and Carpentry & Joinery at Advanced level was done. The analysis was conducted on 2002 to 2004 examinations. Tables 9 and 10 have more details.

Table 9 Cut-scores for Brickwork

	2002	G-points Difference	2003	G-points Difference	2004
Pass	30	-4	26	+4	30
Credit	50	0	50	0	50
Distinction	65	-1	64	+1	65
Sd	15.404		14.597		13.19

Table 10 Cut-scores for Carpentry & Joinery

	2002	G-points Difference	2003	G-points Difference	2004
Pass	30	+5	35	-10	25
Credit	50	+5	55	-5	50
Distinction	65	0	65	0	65
Sd	10.969		10.905		12.00

Discussion of results

As has been established, TD at school level is basic and generalized because it is the foundation stage. However, it would appear that during the first lectures at Craft level, the school level content is covered because not all learners might have learnt TD in school. This is why there is almost no correlation in performance between those who did TD in school and those who did not going by the results of ANOVA. The first three tests in the three Technical colleges support this assertion. The implication of this scenario is that syllabus coverage in Technical colleges delays because a number of students are just at the beginning level of TD. In turn this may affect the motivation of those who did TD at school level since not much challenging work is not presented to them during the first weeks.

Variations in TD content across the three years under study for Machine Woodworking and Vehicle Body Repair could be attributed to non reference of the existing Table of Test Specifications during test development. This can easily distorts assessment standards over the years. The Table of Test Specification indicates that 25% of the total marks of the paper be allocated to Technical Drawing. Again, the possibility is that item setters and moderators did not restrictly adhere to the Table of Test Specifications. In almost all the trades, it seems that a particular question number carries the same topic every year. This could be a sign of a short syllabus. This state of affairs can lead to overtesting of the topics, and hence candidates memorizing or guessing what would come in the examination paper.

Ideally, calculation of confidence intervals on particular cut-score points could help to establish whether or not assessment standards have been maintained over the years. This however was not realistic because of high standard deviations. The high standard deviations reflected diverse performance abilities from one candidate to the other. Probable reasons for this are variations in teacher qualification, high mobility of teachers, inadequate numbers of teachers, laziness in some students, and lack of training of markers, setters and moderators.

Conclusion

Vocational programmes need to be at a higher standard than general educational ones because they are at a higher level of education. However, in Malawi context, such programmes need to start from basics i.e. they should cover what is taught at school level because not every vocational student may have done any foundation course for such a vocational programme. After all, only 13 out of over 700 secondary school centres offer TD.

Apples and oranges?

Comparing standards of vocational and general courses

Stephanie Matseleng Allais, Umalusi

Abstract

This paper makes an argument about the development of a methodology for comparing the relative standards of different courses within the same broad subject areas. The paper argues that different courses can be compared using a clear framework which enables an examination of specified content as well as difficulty levels of examination questions within specified levels of cognitive challenge. It further argues that specified learning outcomes do not seem to be particularly useful in this type of comparison. This is particularly significant in relation to South African education policy that has assumed that outcomes can be the primary vehicle for specifying the level of a qualification or learning programme.

Introduction

The paper analyses research conducted by the author at Umalusi, and published by Umalusi in 2006 as *Apples and Oranges: a comparison of school and college subjects*. The aim of the research was to compare the standards of different courses at the same level. Specifically, the research compared courses between general and vocational education tracks. The paper provides an analysis of the research, and makes its argument based on an examination of the research methodology and a consideration of its implications.

Umalusi is a statutory body with a national mandate to issue certificates at a secondary school level, and monitor standards at both primary and secondary levels. As such, mechanisms for judging the standards of courses are very important to Umalusi. Umalusi awards certificates for learners in schools—the Senior Certificate, or Matric, as well as for learners in colleges—the National Senior Certificate and National Certificates. These two qualifications are supposed to have some degree of equivalence between them, amongst other reasons because they are both at level four of the National Qualifications Framework (NQF). New courses have been developed against unit standards registered on the NQF, which are also supposed to be at the same level. As Umalusi increasingly starts to certify additional qualifications, making judgements about the equivalence between different courses will become an increasingly important part of Umalusi's work.

This paper considers the implications of the research comparing vocational with general qualifications in terms of Umalusi's future work. I argue that this research demonstrates that it is only through a framework which looks at specified content and evaluates difficulty levels of examinations within specified levels of cognitive challenge that comparisons can be made across courses. A further implication of this is that meaningful comparisons can only be made within the same broad subject area. This is at odds with the notion of standards and equivalence that was introduced through the National Qualifications Framework in South Africa.

Thinking about equivalence and standards

Standards and notions of equivalence have become increasingly important internationally, as qualification inflation has caused senior secondary school certificates to loose their relative value, and there are increasing enrolments at higher levels of education systems. A further contributing factor is greater public pressure on governments to be accountable in terms of how taxpayers' money is spent, and as such, education systems are under pressure to prove that they are providing 'value for money'. A particularly important but difficult area is making judgements about the relative standards of vocational and general or academic education programmes.

South Africa is no exception to these trends. There has been much concern about 'standards' in our education system, particularly focused on the school-leaving examination, which in South Africa is the Senior Certificate, as well as on the relative standards of various vocational and occupational programmes. In South Africa, Umalusi is the body with a statutory obligation to monitor the standards of qualifications and curriculum, as well as to moderate examinations at the senior secondary level. However, there are also other quality assurance bodies operating in specific industrial and occupational areas. A National Qualifications Framework (NQF) is supposed to be the mechanism to ensure equivalence and standards across this entire system.

Standards traditionally have to do with levels of difficulty in education systems—the breadth and depth of the curriculum in the different knowledge fields. Breadth and depth are affected by the number of knowledge fields or subject areas that learners study for the school-leaving examination. Examinations have played a central role in standards setting in most education systems, because the level at which learners are assessed has a backwash effect on the level at which they are taught—the cognitive breadth and depth of the exam will influence the cognitive breadth and depth of the implemented curriculum. During the process of marking examinations, teachers become familiar with what the expected 'standard' is—they get a sense of what is meant by a pass in a different subject area, what is meant by a high achievement, and so on. This then feeds back into the classroom and has a strong influence on teaching in the final phase of schooling.

However, in South Africa the NQF as well as the outcomes-based education curriculum that was developed for the school system were supposed to be significant mechanisms for solving the problem of 'standards'. They were both based on the idea that the specification of learning outcomes is a useful mechanism for determining the cognitive level of a qualification or learning programme, and for ensuring and determining equivalence between

different courses within different subject areas. Each of level of the NQF is defined by 'level descriptors' which are supposed to 'describe the nature of learning achievement, its complexity and relative demand at each level of the NQF, distinguishing between the learning demands at each level' (SAQA 2001, p11). Learning outcomes for each qualification and programme are supposed to be designed against these level descriptors. The learning outcomes, often referred to as 'standards' are in turn supposed to ensure that both learners and people who are assessing learners have a clear description of 'what must be assessed, in what contexts, and the standard of performance required' as well as a 'a means of recognising achievements (records of learning and/or competence portfolios will indicate what qualifications and unit standards have been achieved by learners)' (SAQA 2000, p. 17). Similarly, in the outcomes-based curriculum that was developed in the school system, was premised on the notion that if learning outcomes were specified, the curriculum could be developed against them, and this would ensure a clear sense of what it was that learners were supposed to learn, as well as what they were competent to do after assessment.

In particular, one of the aims of the NQF was to raise the status of vocational education in society, by designating vocational qualifications at the same level as general qualifications. Many countries internationally have used qualification reform as a mechanism to try to improve standards, as well as to clarify the relationship between different qualifications at different levels, and, as in South Africa, one important feature of qualification reform at senior secondary levels internationally has been attempts to create equivalence between vocational and academic qualifications. Vocational education is of increasing concern to governments internationally, and the South African government is no exception (Allais 2006). However, in many countries around the world, resolving the relationship between vocational and general education programmes at the senior secondary level has proved fraught with difficulty. Nonetheless, it is increasingly important, if the government is going to achieve its aim of encouraging significantly higher numbers of learners to enrol in vocational programmes. Thus, partly because of the increasing pressure for goverments to expand enrolments in vocational education, as well as because of increasing numbers in secondary schooling, notions of 'standards' and of 'equivalence' have increasingly been under the spotlight.

In the context of the pressure for maintaining and improving standards, as well as providing accurate information about equivalence, research was designed to assist Umalusi in understanding the practical reality in various courses currently on offer in South Africa. The research therefore involved an in-depth look at some of the actual courses offered in general and vocational education programmes. The research was not premised on the idea that equivalence is necessary or desirable, simply that it has been an important policy goal, and that formally, and from the point of view of information provided to learners, the qualifications currently are supposed to have a correspondence.

It is important to note that the research could be seen as badly timed, as both the current Senior Certificate and the current vocational programmes were about to be phased out as the research was implemented. In addition, the current vocational programmes have been the subject of much criticism, and are generally perceived to be of an inadequate standard—it could be questioned, therefore, why such research was necessary. Umalusi felt that, notwithstanding imminent policy change, serious questions and debates will remain about

the nature of curriculum change which is needed, particularly in vocational education. A substantive understanding of the exact nature of these courses and potential problems with them will be a significant contribution to future policy reforms. But more importantly for Umalusi's future work and for education policy in South Africa, the research offered possibilities for thinking about standards, and it is this aspect of the research that is the focus of this paper.

A particular concern in this regard was Umalusi's relationship with other quality assurance bodies. Umalusi currently quality assures examinations, and accredits private providers as well as examination bodies, in order to be able to issue certificates for specific qualifications. However, Umalusi is expected to co-operate with quality assurance bodies which are constituted under the Sector Education and Training Authorities (Setas), which operate in a completely different quality assurance model, based on delegated assessment as opposed to examinations. Umalusi is also responsible for monitoring the suitability of standards for the qualifications that it certifies in the General and Further Education and Training Bands of the NQF. As such, the research also aimed to assist Umalusi to be able to engage more meaningfully with debates about pathways within senior secondary education. In short, the research had considerably broader aims than merely making pronouncements on the relative standards of the selected subjects: Umalusi wanted to reflect on its current quality assurance mechanisms, and to develop meaningful ways of co-operating with other quality assurance bodies.

About the research project

The broad aims

This section of the paper provides a very brief overview of the research that Umalusi conducted, including a very brief discussion of the research findings, and some of the conclusions and recommendations.

The research was specifically designed to develop a better understanding of the current nature of vocational courses in South Africa—to determine whether they are in fact equivalent to school subjects at the same level and whether they prepare learners for higher education.

Preparation for higher education was included partly because linkages with higher education is an important way of improving standards of senior secondary programmes. Also, input from higher education institutions can also provide valuable information about the relative standards of different courses. In the United Kingdom, creating vocational qualifications that lead to higher education has played a significant role in encouraging learners to enrol for vocational programmes (Wolf, 2002). Countries with strong vocational programmes at secondary level, such as Germany, Austria, and the Scandinavian countries, tend to have established pathways into higher education programmes within the vocational area being studied. However, in South Africa learners studying in FET colleges have generally not intended to go to university, and most of the courses were not designed with higher education in mind. The courses were also often designed to cater for learners who had not succeeded in school, which made it even less likely that they would be designed from the point of view of higher education study.

Although preparation for the workplace is a crucial aim of vocational qualifications, it was not possible to include this aspect in this research; it will be the focus of a follow-up research project.

The qualifications in question

There are three main types of certificates that Umalusi issues: the Senior Certificate, for learners in school, and the National Senior Certificate and National Certificates for learners in colleges. In the past in South Africa subjects for Senior Certificate, commonly known as the Matric, were offered on two main levels: Higher Grade which was supposed to be more cognitively challenging, and Standard Grade which was designed as an easier alternative. College or vocational qualifications were seen as the equivalent of the school qualification, but at Standard Grade level. Under the NQF, there is a series of qualifications all designated to be at level four—the level of the Senior Certificate. Various new qualifications and part qualifications (unit standards) have been registered at this level. Currently, these are not issued by Umalusi, although there is pressure on Umalusi to co-operate with other quality assurance bodies with regard to some parts of these qualifications.

For the purposes of the research, a few individual subjects that are taken in the different learning pathways were selected. The subjects selected were Mathematics, Science, and English courses, as well as courses within the broad area of Hospitality. There were three different kinds of courses selected. The first were those taken by learners in high schools. The second were those taken by learners in Further Education and Training colleges—in other words, learners enrolled in vocational programmes. The third type of courses were new courses that had been developed against unit standards registered on the NQF. The intention of the research was to compare courses in the *same* subject areas, as well as to explore the extent to which the different courses prepared learners for higher education. (Note that this is a far more modest investigation than the notion of equivalence introduced through the NQF, which makes a claim across subject areas.)

It is worth noting that research was inevitably incomplete because each course is part of a total package. The location of the course within this package affects what the course is—that is, the role that the course plays in the development of the learner. The standard of the qualification obtained is a product of the combination of the different subjects taken; the subjects should not be looked at only in their own right. The vast majority of learners writing the Senior Certificate are studying six subjects at the same time. In general these subjects are part of a three-year programme—what is examined in the final examination is taught over the second and third years, with the first year in most cases laying the basis for the next ones. Learners writing N3 Engineering in colleges, on the other hand, may study for only one year in total, taking four subjects done in a series of three trimesters each (for N1, N2, and N3 respectively). The final examination for N3 is based on the last trimester only. Learners studying for the National Senior Certificate usually do six subjects over two years; the final examination is often based on the Grade 11 syllabus as well as the final year or Grade 12 syllabus, and sometimes even tests items taught in Grade 10. In some subjects only the Grade 12 syllabus is explicitly tested. Obviously, for the National Senior Certificate, the National Certificate, and the Senior Certificate, the knowledge tested in the final year is cumulative, and builds on knowledge and abilities taught in earlier courses. From the point of view of preparation for higher education, this is particularly important. It could be

possible, for example, that a course which provides training in memorization and organization of knowledge in an area like hospitality, without any conceptual or abstract thinking, could play some role in preparing learners for higher education, as long as the other courses that learners are taking develop the other skills and abilities required. It is thus difficult to make a judgement about a course on its own, in terms of higher education preparation.

The research methodology

Two teams of researchers commissioned by Umalusi examined courses in the different subject groups in order to provide answers to the two research questions. First, a team of four practitioner evaluators for each subject evaluated syllabuses, examination papers, memoranda, and marked scripts for the courses, as well as course packs for some of the new NQF courses. Each team consisted of two Grade 12 educators and two college lecturers who had produced consistently good results. One member of each group had participated in Umalusi's 2004 *Investigation into the Senior Certificate Examination*. Second, a team of four higher education experts, drawn from universities and universities of technology, evaluated the same data, as well as the reports of the practitioners.

Evaluators were asked to analyze the courses using three main categories of comparison: content coverage, key concepts and procedures, and expected outcomes. Criteria and guidelines within each category were developed, and evaluators were asked to rate examination questions on a scale of cognitive challenge. The three categories and the scales of cognitive challenge were used as the basis for judgements about both research questions—the equivalence of the courses, and the extent to which they prepare learners for higher education.

Evaluators worked jointly in workshops and individually to produce analyses of their subjects. A single report was produced by each of the two groups of evaluators in each subject area. Umalusi produced this report on the basis of a synthesis and analysis of the subject reports.

Overview of findings

The evaluators found that, based on the amount of knowledge to be learned and the level of cognitive challenge at which it is examined in the courses, the college courses are seriously lacking. They further found that the unit standards-based courses could not be evaluated.

In Science, Mathematics, and English (Home and Additional Language) there was clearly no equivalence between the school and college subjects—the National Senior Certificate courses were seen as far less substantive and were tested through examinations which contain few challenging questions. The school or Senior Certificate subjects were found to be broader than the corresponding college or National Senior Certificate subjects, and the examinations for the former certificate were in general considerably more challenging. It was argued that vocational question papers needed to be more challenging and less predictable. In other words, to again work with the metaphor of the title, what was found was that the courses were not apples and oranges, but more like apples and half apples. Apples and oranges would imply that the courses are different in appropriate ways and good in their

own terms. What seems to be the case, however, is that, particularly for Mathematics, Science, and English, the college subjects are weaker versions of the school subjects.

There are no formally equivalent courses in Hospitality across schools and colleges, although there is a lot of common content in the courses. However, it was most difficult to conduct a meaningful comparison, and in fact only part of the two courses was compared.

None of the four college subjects examined as part of this research were found to prepare learners for degree study in higher education. However, many of the school subjects were also seen as seriously lacking in the content, skills, and levels of cognitive challenge required to prepare learners for degree study.

Some of the findings were relatively straightforward. For example, for both Mathematics and Science, the college courses are taught over a dramatically shorter period than the Senior Certificate courses, providing learners with insufficient time to master the content in question. This correlated with a general finding that the college courses covered the content in less depth. Further, the Science and Mathematics courses in college were examined through a single question paper, while the school subjects had two. The college courses had a more rigid format, and this, evaluators argued, was likely to lead to more predictable examinations over time. Evaluators also pointed out aspects such as the speed at which learners were expected to work in the different papers, arguing on this basis, for example, that the college Mathematics paper was easier in this respect than the Standard Grade school paper. The time problems in some of the vocational courses is a historical legacy of the trimester system, which was developed for apprentices who would also have been learning in the workplace. It was argued that this system was clearly no longer appropriate, and in particular was seen as limited in terms of learners' chances of mastering a substantive body of knowledge.

Differences in terms of content were more debatable. For example, with regard to Science the college and school courses are relatively similar in terms of the *amount* of content specified, although, due to the shorter time of the college course it seems less likely that learners would be able to master the content, and more likely that topics will be covered in less depth. However, there is a substantial difference is in the nature of prescribed content. The college course consists of Physics only, while the school course includes Physics and Chemistry. The former course does have additional Physics topics and industrial applications not covered in the Senior Certificate course. Some of the evaluators felt that the courses could be seen as different but equal. Others strongly argued that the inclusion of Chemistry in the Senior Certificate courses make them substantially broader, and of a higher standard because different levels of scientific thinking (macro and micro) were being taught.

Similarly, in English there was considerable debate about literature. Some evaluators felt that the exclusion of literature made it very unlikely that learners would have sufficient exposure to extended reading of continuous prose, and as such they would be unlikely to become very good readers. The argument was that literature is the best (or easiest) way of ensuring that learners master complex language forms. Others argued that literature was not important, and that any kind of continuous reading should be seen as adequate. The research itself did

not shed light on this question. In the courses under examination, it was only the courses that taught literature that had anything approaching sufficient amounts of reading. Nonetheless, evaluators still maintained that while the other courses in the present study were seriously lacking in terms of reading, a conclusion should not be drawn that literature per se is necessary in an English course.

Another debate in English related to breadth relative to depth. The evaluators found that with regards to writing, the students on the college programme did more writing, but of a far more limited variety.

Evaluators made various comments on the form and presentation of the syllabuses. For example, the Mathematics evaluators felt that both sets of syllabuses were inappropriately compartmentalized. In both English and Hospitality the syllabuses were found to be in a state of some disarray.

Considering the methodology with regards to making judgements about standards

The methodology of this research raises particular points of interest for a discussion about standards and about equivalence. Making judgements about standards and the difficulty level, as well as appropriateness of curricula and examinations, is fraught with difficulty. Nonetheless, benchmarking the breadth and depth of curriculum and assessment is the task that Umalusi is mandated to carry out in South Africa. There were a few key components to the methodology of this research project which could have bearing for future work.

It is common cause that subject experts should make judgements on these kind of questions. However, this particular methodology goes further to guide the nature of the judgements being made. The evaluators were asked to use three different lenses to investigate the different curricula: content, concepts, and outcomes. The intention was to compare all three of these for the intended curriculum (the prescribed syllabus) as well as for the examined curriculum (the examination question papers or assessment tasks). Within these broad categories, evaluators were provided with guidelines and further specifications.

It is important to consider how three different categories were able to assist the evaluators to make their judgements.

Firstly, consider the third category: outcomes. As discussed above, the specification of learning outcomes was introduced in South Africa as a mechanism to ensure that courses were of the required standards, as well as to enable comparisons between courses. This research raised interesting problems with this notion. Consider the following problem faced by the evaluators: they were confronted with a set of courses developed against learning outcomes registered on the NQF. In some instances, different courses claimed to lead to the same learning outcomes. How were they to judge whether or not these courses achieved the stated outcomes?

Bear in mind that the only thing prescribed, in this model, is the learning outcome. There is no prescribed content. The 'courses' that the evaluators were confronted with were in fact a collection of suggested classroom activities. But how, then, could evaluators make a judgement about the relative standard of different programmes that both claimed to lead to the same outcomes, and consisted of very different suggested activities? How, in the absence of summative assessment instruments, could evaluators make any sense of what level of knowledge and skill was expected from learners? Contrary to claims of outcomes-based qualifications systems, the outcomes themselves did not contain any way of resolving this problem—it was clear from the different courses that the same outcome could be interpreted in very different ways. Because of this problem, the evaluators in fact were not able to conduct a systematic evaluation of these courses.

A few possible concerns were raised. For example, the Mathematics evaluators felt that it was possible that because the contextualized Mathematical Literacy course was so embedded in the context there was a serious danger of losing sight of the mathematical competencies being built. The English evaluators felt that the suggested texts in the course packs were very inadequate. However, what is more significant was the fact that the evaluators were unable to compare these courses, because this implies that an approach to curriculum design in which only outcomes are specified will make quality assurance unreliable or untenably expensive. If quality assurance bodies are to make judgements about courses in which there is no stipulated content and no summative assessment, they will have to physically visit classrooms, moderate individual assessments, and evaluate individual learning programmes. All of this sounds nice educationally, but is completely untenable from a financial point of view in a large education system, particularly in a poor country. Nonetheless, this is the mechanism adopted by many of the other quality assurance bodies (operating under the Sectoral Education and Training Authorities, or SETAs) with which Umalusi must work.

The notion of aims/objectives/outcomes when considered in relation to prescribed content, as part of a syllabus document, proved to be a more useful notion. A consideration of what the course was aiming to do—such as, preparing learners for the workplace, as opposed to preparing them more generally for further education and taking their place in society, enabled fruitful debates and discussions about what kinds of content areas should be prescribed. In other words, prescribed outcomes or aims are a very important component of a syllabus, but are not useful when considered separately from a syllabus.

If outcomes did not enable a meaningful comparison, how was the majority of the evaluation conducted? Through developing tables that outlined the topics covered in the different courses, evaluators were able to create a picture of the relative breadth and depth of the courses. They were able to demonstrate that an examination of the nature and number of topics specified in the syllabuses was a significant mechanism to make judgements about the relative standards of courses. This does not mean that a specification of the key content areas eradicated the need for debate and expert judgement—as discussed above, there was considerable debate, for example, about the replacement of Chemistry with additional topics from Physics in the Engineering Science course, and about the inclusion of literature in an English course. However, a grid of topics covered enabled a first level of comparison, and also provided the framework for more evaluative discussions and judgements to take place.

However, by only comparing the prescribed content, evaluators were not able to make a judgement about how challenging the course was, what kinds of cognitive abilities were actually expected from learners, and at what levels of difficulty. It was only a consideration of the question papers that enabled this kind of judgement. Evaluators developed tables specifying different levels of cognitive challenge, and providing examples within their subject area. As an example, the table below shows the matrix developed by the Science evaluators.

Table 1: Categories and levels of cognitive demand within Science exams

Category	Level	Descriptions	Examples
Factual recall / rote	**Simple (1)**	State a simple law or equation	State Newton's laws etc.
	Medium (2)	Recall complex content	Process for lab preparation of chemical compounds; testing for presence of diff chemicals; inorganic chemical interactions
Understanding of concept / principle	**Simple (1)**	Simple relationships; simple explanations	Relationship between resultant and equilibrant; explain what is meant by … ;
	Medium (2)	Counter-intuitive relationships; Qualitative proportional reasoning; more complex relationships or explanations	Direction of acceleration for free-fall; effects of changes in circuits; identifying acid-base conjugates, redox pairs etc; simple influences on dynamic equilibrium
	Challenging (3)	Identify principles which apply in a novel context	Identify all influences on realistic motion; identify isomers of organic compounds; complex influences on dynamic equilibrium
Problem solving	**Simple (1)**	Simple procedure; plug into formula with only one unknown; no extraneous information; known or practiced context	Given current and resistance, calculate voltage; etc
	Medium (2)	Construction or interpretation of diagrams; problems with 2 or more steps; basic logic leaps; proportional reasoning; interpretation of table of data	Graphs of motion; force or vector diagrams; concentration or molar calculations; naming of organic compounds; writing and balancing equations for reactions
	Challenging (3)	Complex abstract representation; combination of concepts across sub-fields; complex problems involving insight and logic-leaps; formulating new equations (using all unknowns); problem solving in novel context	Interpret complex graphs; translate between various graphs of motion; combine equations for mechanical energy and motion; combine gravitational and electrostatic forces; complex circuit calculations; combination of various factors influencing equilibrium

Note that the tool includes different kinds of cognitive processes that are important for Science. These are divided into different levels of difficulty, with a description of what exactly is meant, and examples. Systematically applying tools like this one to items in the question papers enabled evaluators to make judgements about the relative standards of the different papers. Again, the tool did not remove the need for debate and discussion—there will always be disagreement amongst subject experts about difficulty levels, and a range of other factors affect how difficult any particular paper is for individual learners. Factors such as predictability of question papers is obviously a major concern here. However, the tools developed enabled evaluators to make informed decisions that were at the same time more

accountable, from the point of view of the research (or, the quality assurance body) than simply making statements based on expertise.

Implications for Umalusi's work

In sum, the research provided a much-needed insight into an issue which has been dealt with in overly general terms in South African education policy—the notion of equivalence. It became very clear during the course of the research that it is very difficult to talk about equivalence in generic terms—specific content areas and tested levels of cognitive challenge within content areas need to be investigated. The evaluators found, however, that it was quite possible to compare the different courses within the same broad subject area when a clear framework which looked at specified content and evaluated the difficulty level of examinations within specified levels of cognitive challenge was used. To extend the metaphor of the title, apples and oranges *can* be compared from the point of view of nutritive value—composition and amount of various dietary elements.

For the purpose of future considerations about standards, and the development of systems and processes to deal with standards and equivalence, one of the important outcomes of the research was that it showed clearly the problems of using only outcome statements as an expression of desired standards for a course. Stipulating only learning outcomes is not a viable approach to curriculum if any degree of standardization or equivalence is believed to be important. Dramatically different courses claimed to lead to the same or similar outcomes, and the outcomes themselves did not contain anything which could resolve this problem. As stated above, it was the evaluative framework, which consisted of comparing the breadth and depth of specified content as well as judgements about assessment instruments, that enabled any real understanding of the courses. The specification of outcomes is like the orange peel—there is nothing substantive that can be seen or judged from the outcomes alone. This has serious implications for Umalusi when it considers how to work with other quality assurance agencies, who may base their quality assurance processes on evaluations against learning outcomes.

Probably the strongest message that came from the research was that urgent attention needs to be paid to curriculum development in South Africa—particularly the development of user-friendly syllabuses which provide clear indications to teachers about the key knowledge areas to be covered and the levels of cognitive challenge to be assessed. Syllabuses should also guide teachers with regard to teaching approaches for the key content areas; this was particularly urgent in vocational programmes.

Following from this, what will be essential for Umalusi in terms of its role in monitoring the adequacy and suitability of standards and qualifications is the consolidation of the categories and criteria used in this research into tools which can be used to make judgements about curriculum statements, syllabuses, and examinations. The categories, criteria, and evaluative scales developed in this research and the 2004 *Investigation into the Senior Certificate Examination* will provide a substantive basis from which Umalusi can develop tools for the evaluation of syllabuses and examinations. From a systemic point of view, Umalusi will also have to engage with other quality assurance bodies to attempt to build a common model.

References

- SAQA (2000) The National Qualifications Framework and Standards Setting. Pretoria: SAQA
- Umalusi (2006) *Apples and Oranges? A comparison of school and college subjects*. Council for Quality Assurance in General and Further Education and Training. Pretoria.

Determining standards in practical subjects

The Zimbabwean experience

Betty Mutambanengwe, Zimbabwe School Examinations Council

Abstract

Standards in educational assessment are set to ensure that learning goes on in the classroom and they can only be raised if teachers and assessors do their work effectively. In Zimbabwe, the general public, due to ignorance of the processes of assessment, feels that localization of public examinations coupled with paucity of resources have led to declining of standards. But with the experience gained over the years from the localization of O-Level examinations in the mid- 1980s, the Zimbabwe School Examinations Council is rated among the best in the region when it comes to assessment standards (NARIC, 2002).

This research was carried out to determine how stakeholders view standards of assessment in practical subjects in Zimbabwean schools taking into account the use of locally available resources, during the teaching and learning process.

Direct observation, interview and questionnaire techniques were used to gather data from schools in urban, rural, day, boarding, mission, and private settings including Food and Nutrition item writers and markers, Hospitality Industry Human Resources Personnel and ZIMSEC employees. The data revealed that the use of locally available resources in most schools has led to a lot of improvement, which has made the learning of practical subjects meaningful.

Introduction

The standards that a nation strives to achieve are embedded in the national curriculum, which in Zimbabwe, is developed by the arm of the Ministry of Education Sport and Culture—the Curriculum Development Unit (CDU). The Ministry plays a major role in the initial stages of determining the curriculum to enable the Quality Assurance Division to monitor and ensure that standards are maintained in schools. The Zimbabwe School Examinations Council (ZIMSEC), as the examining board, is involved in all subject panel meetings for the main purpose of identifying areas that can be examined (the assessment

scheme). Thus the responsibility of ZIMSEC in the education system is mainly that of measuring learners' achievement/performance standards and the certification of candidates' performance at national level.

This paper focuses on how standards are determined and maintained in practical subjects at ZIMSEC with special reference to Food and Nutrition at all examination levels. Also looked at are the tasks of practical subjects assessors, how standards are linked to accountability, keys to quality assessment, and findings from respondents on their views about assessment standards at ZIMSEC. Recommendations as a measure to guard against lowering of standards are given.

Before the study was undertaken to determine how stakeholders view the standards of assessment particularly in practical subjects, it was assumed that the use of locally available resources in the teaching and learning is likely to produce sub-standard products and that in the absence of universal resources, the same skills cannot be mastered. Hence the study was carried out to find answers to the following questions:

- Do students from different centers acquire the same skills and concepts irrespective of their resources and environment?
- Is the final grade awarded to one candidate comparable to the same grade awarded to another candidate in a different environment and different practical subject?
- Should there be different marking schemes during marking to cater for candidates with different learning backgrounds?
- What attributes do employers consider when recruiting school graduates?

Theoretical framework

In this study literature was reviewed in order to put into perspective what was encountered by other researchers in the subject of educational and assessment standards and relate it to Zimbabwean experiences in assessment standards in practical subjects like Food and Nutrition, where resources vary considerably from school to school.

Education and examination standards

The concept of standards has been interpreted in a variety of ways in different education systems at any particular time (Goldstein and Heath, 2000). In that vein, Gipps (1990) has used the word standards in education in three distinct ways, which are: **attainment, levels of educational provision, and matters of conduct and social behaviour.** In other words, standards logically provide the foundation for testing, and testing results are a measure of both learner and teacher success. A standard is a criterion that examinees need to achieve in order to be deemed competent. Hence, for practical subjects in Zimbabwe, a candidate is awarded a certificate after going through a course and proving to examiners and the system as a whole that he/she is capable of performing and/or producing knowledge rather than reproducing information. This therefore implies that in order to maintain assessment standards in practicals, emphasis should be on the **doing** (what candidates are able to do) and the **knowing** qualities of an individual. The major challenge is that examinations as a system ought to be such that examinees are given meaningful tasks, which enable them to make use of available resources to solve problems on hand.

Determining standards

Establishing clear, rigorous standards that specify what learners should know and be able to do is very crucial in the education and assessment of candidate performance. It is now evident that due to globalization and increased international competitiveness, governments are increasingly recognizing the importance of the standards of academic and professional qualifications. In other words, the education system is undergoing significant changes in many parts of the world, with the aim of improving the quality of education and its suitability for each and every learner. This implies that no child should be denied access to education. Therefore, the assessment system of a country ought to be appropriate for learners of all abilities while still maintaining comparable standards. In defining standards, Cresswell (1996) points out that there are critical areas on which value judgements must be made. The areas are: **what** attainments are assessed and the **quality** of the observed performance during which those performances are demonstrated. **What** is the content upon which judgements are passed by experts and the **quality is** skills, which the question papers tests. Therefore, in defining examination standards we should always involve a consideration of the value attached to the knowledge, skills, and experiences which are assessed and then focus on thoroughness and excellence.

In Zimbabwe, because the general public is ignorant of the assessment process, they often express in the media the feeling that the localization of public examinations coupled with paucity of resources, have led to the declining of standards. They assess the success and failure of schools by how well candidates perform on standardized tests and eventually in the world of work. But with experience in assessment gained over the years from localization of O-level examinations in the mid-1980s, ZIMSEC is rated among the best in the region when it comes to assessment standards (UK NARIC, 2002). On the other hand, feedback from prospective employers such as the Hospitality Industry, in the case of Food and Nutrition graduates, helps to allay fears that localization has led to lowering of standards.

During syllabus development, CDU involves personnel from industry and commerce (the business community) together with representatives from the education system (schools, technical colleges and universities) so that they influence the kind of expected competences to be developed in schools as an innovation meant to discourage rote learning in practical subjects like Food and Nutrition. When it comes to moderation of the marking process, the Subject Manager and National Chief Examiner (NCE) set the standards on what is expected of candidates and how marks should be awarded during the marking process as markers go out to marking centers. The NCE's marking is the standard nationally. After marking there is transcription checking then grading and grade reviewing to bring the marking in line with set assessment standards.

Sanders and Horn (1995) support what ZIMSEC views as essential item development practices, where the draft items are subjected to rigorous validation, reliability testing and standardization procedures, which include moderating, proof reading, vetting, and pre-testing. As standard setters, during item development, panelists agree upon what to include in the marking guide and the weighting (mark allocation). Thus they develop quality assessment tools, which discriminate good candidates from poor ones. This makes assessment authentic and has proved useful in comparing, generalizing, and indicating levels

of attainment based on set standards, which call for higher order reasoning skills and academic growth of the learners within their local environment.

Methodology

In this survey, direct observation, interview and questionnaire techniques were used to gather data from schools in urban, rural, day, boarding, mission and private settings. Included in the sample were Food and Nutrition item writers, markers and teachers, Hospitality Industry Human Resources Personnel, and ZIMSEC employees.

An analysis of national pass rates in Food and Nutrition at O-level since localization of the examinations revealed that the involvement of all these experts during syllabus and item development means that learners are now going through programmes, and are tested on real-life experiences. The candidates are given an opportunity to demonstrate their abilities using locally available resources. The table below indicates the trend of entries and results of Food and Nutrition at O-level over a period of four years interval from 1988 to 2005.

Table 1: Progression of entries and pass rates in food and nutrition at O-level

Year	Entry	Candidates Who Obtained Grade C or Better	Pass Rates
1988	3 815	1 620	42.5%
1993	4 217	2 333	55.3%
1997	5 638	3 549	61.2%
2001	6 526	4 099	62.8%
2005	5 948	2 769	46.6%

Statistics show a gradual progression of candidature and pass rates. However, the 2005 statistics might have had other factors such as economic hardships due to drought, relocation of families, teacher mobility, and so on, that intervened during the learning process and examinations. Otherwise the level of difficulty of question papers and quality of candidates has been more or less the same according to comments from the NCEs and Subject Managers during the grading meetings of the different examination sessions.

In essence, it implies that assessment agencies should renew their commitment to the standards required of high quality assessment so that at the end of the day they can say, with confidence, that what is offered for the purpose of certification is a fair and equitable measure of all candidates' performance.

The question now is 'how can we judge whether a particular assessment method is appropriate for practical subjects?' For fitness of purpose, when setting and designing assessment tasks, decisions are made about what is important and workable with the hope that this may not involve difficult compromises which are likely to lower standards.

Tasks of assessors of practical subjects

Since emphasis is on doing and knowing, and maintaining standards, assessors are tasked to check on:

- the functional aspect of a product;

- accuracy of measurements, for instance, weighing of ingredients in Food and Nutrition, or lengths of pieces of wood in Woodwork etc;
- the aesthetic effect the product has on the assessor (presentation or impression);
- available resources (human, material and financial) depending on the environment.

From these tasks, it can be observed that the major area of concern is the source of variation in standards emanating from the aesthetic effect, which is highly subjective as we all have different ways of viewing things depending on our backgrounds. Assessment of this nature can be problematic in that it brings in a lot of outside factors, which have a bearing on marker performance. At ZIMSEC this is one of the areas of emphasis during co-ordination meetings before marking, to ensure that standards are maintained.

Ewell (2002) points out that learners' competences should reflect critical thinking, communication skills, quantitative literacy and problem solving skills (real-world applications in the workplace and other settings). Generally, today the education system is aiming at developing an individual who is capable of making complex and wise choices, and reasoned judgements that affect the individual's well-being because the 21st Century workplace demands sophisticated problem-solving skills. We say standards are comparable if a learner is able to perform in a changing environment and participate fully in society applying learnt skills, which include: communication, application of numbers and use of information technology, to solve problems in a given situation (SCANS, 1991).

How to maintain set assessment standards

In Zimbabwe, during training and co-ordination meetings, markers are instructed to guard against marking down candidates who are creative and those who show originality in their projects as we strive to run away from solely relying on textbooks. Markers should avoid biased assessment. To get round the problem, at ZIMSEC, practical subject markers are recruited from the same provinces as the candidates, where they share the same environmental experiences and its resources. These examiners rotate schools so that one marker does not mark the same school year after year. Also during co-ordination meetings, assessors are tasked to measure those skills and abilities, which have been established as being important and those, which the course sought to develop. All markers for a specific subject component use the same marking scheme agreed upon during the national co-ordination meeting so as to maintain standards.

As examiners go out to examination centers to mark practical subjects like Food and Nutrition, Building Studies, etc, they are provided with a guideline with clearly stated purposes of assessing practical work. Some of the purposes include that practical work should be assessed to:

- reinforce theoretical teaching;
- develop an understanding of the subject;
- demonstrate practical skills;
- reinforce the concept of economical use of resources (resource scarcity); and
- develop an understanding of safety precautions in the workshop and at home while handling machinery and other tools.

Therefore, since practical work involves **process** and **product** (Jones and Bray, 1986), the process being the more difficult of the two, examiners undergo rigorous training in assessing the processes as they occur in practical subjects. At ZIMSEC, in Food and Nutrition in particular, as part of their training, examiners go to a specific school where some students have been tasked to carry out a mock examination. Trainee examiners mark in groups concurrently. At the end of the session, they compare results and discuss differences basing on the National Chief Examiner's standards. The same process is carried out during live marking of practical Food and Nutrition, the difference being that multiple marking is done by the National/Regional Chief Examiner/Team Leader and the assistant examiner during moderation. The assistant examiner gets feedback on possible causes of deviations if any and make necessary adjustments to his/her marking scheme in line with the leadership's standards.

As a measure against subjective marking and to avoid biases, guidelines on what skills to look for during practical examinations are given to examiners during co-ordination meetings, which include:

- Comprehension of the process or task;
- Awareness of appropriate contexts;
- Knowledge and understanding of a range of approaches;
- Understanding of bases for decision-making;
- Ability to carry out practical tasks with skill;
- Ability to integrate theory with practice;
- Ability to address issues arising from the performance context or situation;
- Creativity and originality;
- Ability to work sensibly, responsibly and supportively with others as candidates share equipment in most schools.

Standards versus accountability

While in schools teachers are held accountable for the success and failure of students, examination boards are also, on the other hand, accountable for authentic assessment, raising and maintaining standards with the blessings of the national and state policy makers (Black and Wiliam, 1998). Harman (2002) states that practical subjects, while they emphasize more on the doing than just knowing, when it comes to assessment, the same rules for assessment of educational quality and standards are applied. Quality education is a universal goal, while authentic assessment ensures that learning is driven by what teachers and learners do in the classroom and not examination driven.

Therefore, standards ought to be accompanied by accountability. However, it should be noted that in practical subjects, the examination system relies heavily on human decision-making, human abilities, and human behaviour. Generally, there is a mistrust of human judgement in assessment particularly in subjects like Food and Nutrition whose assessment is highly subjective, and where there are likely to be disagreements which undermine the impartiality of assessment. However, the key issue in resolving disagreements is recognizing each other's professionalism and offering comments in a professional manner.

Keys to quality student assessment

Quality assessments arise from clear and appropriate student learning targets spelt out in the curriculum, content standards, benchmarks, learning objectives or outcomes, learning goals, or essential academic learning requirements. They serve a focused and appropriate purpose such as why targets are being assessed, who will use the results and for what purpose (e.g. recruitment into the Hospitality Industry). Purpose affects how one assesses. Quality assessments use appropriate methods of assessment, which accurately reflect the achievement targets and purposes. They eliminate possible sources of bias and distortion for interested parties to be confident that results really reflect what a candidate knows and can do (issue of schools on the shop-floor/attachment). Quality assessment ensures reliability, validity, fairness and equity of the assessment instrument.

Presentation and discussion of the results

Stakeholders' views of assessment standards in the education system

The researcher, while on a different mission in schools during examinations, observed that during practical examinations in Food and Nutrition some candidates prepared interesting and skillful traditional dishes at the same time answering the questions as they substituted some ingredients stated in recipe books with foods from their local environment. It was also observed that the use of local resources develops that love of learning, curiosity, interest, intelligence, and creativity.

Responses from examiners and teachers were that schools which are well equipped tend to have an unfair advantage over those in poor communities as this has an influence on opinions of external assessors. The aesthetic effect of products and environment has a bearing on the final score a candidate is likely to get. However, during co-ordination meetings, examiners are instructed to guard against such biases as it should be the other way-round because for someone who is battling to come up with something presentable when basic equipment is improvised, that individual deserves a reward for being innovative. Under such circumstances, in order to maintain standards, at ZIMSEC special consideration is only taken into account in extreme cases such as power cuts in the middle of a practical examination. However, comments from teachers were that ZIMSEC is now experiencing higher pass rates as compared to the past partly due to coaching from teaching examiners as well as the use of locally available resources. Bringing lessons closer home makes teaching and learning more meaningful without compromising standards. Siyakwazi (1997) argues that the curriculum and its assessment should have a local bearing for relevance of what transpires in the classroom. On the other hand, learning becomes more challenging and exciting as learners are forced to be innovative while exploring their environment in order to get solutions to problems.

Responses from the hospitality industry personnel were that they prefer to recruit school products who have background knowledge of food preparation as they are easy to train. Currently, the use of locally available resources has made traditional dishes popular in hotels, restaurants and other food outlets in Zimbabwe where school graduates' experiences in the use of local resources is highly appreciated. ZIMSEC employees interviewed who have a dual social role of being assessors and parents of the candidates assessed, indicated that they also support the use of local resources, which partly addresses the socio-economic problems.

The results in general seem to point to the fact that the use of locally available resources has had a positive effect on the teaching, learning, and assessment of candidates as evidenced by the increase in pass rates and entries in Food and Nutrition in particular.

Conclusion

Practical subjects assessors the world over are reminded that the main aim of offering practical subjects in schools is to stimulate innovativeness and creativity among learners so as to make learning and assessment more interesting and meaningful. In essence, the assessment instrument we use and the results should reflect a well-rounded individual who is able to apply knowledge by reasoning and solving novel problems. Prospective employers, on the other hand, expressed that they prefer prospective employees to have not only good communication skills but manipulative skills as well for the individuals to fit in industry.

From the general consensus of the respondents in the survey, ZIMSEC has established credibility in the market by interacting and standardizing with other examinations boards, and achieving international recognition through students' acceptances in universities outside our borders. On the other hand, the use of locally available resources has partly addressed the socio-economic factors affecting most developing countries like Zimbabwe. However, the biggest challenges for ZIMSEC are to continuously provide the market with evidence of its consistency in the marking and awarding of authentic scores in all subjects across the curriculum and to uphold its integrity as well as ensure stakeholders that anything indigenous does not mean that it is substandard. It is a real challenge to achieve quality and maintain standards across the curriculum especially in a situation where learning resources vary considerably as is experienced in Zimbabwe.

Recommendations

In the light of the research findings, the following recommendations are made:

- It is recommended that policy makers, curriculum designers, and examination authorities work hand-in-glove for the education systems to be meaningful and relevant to stakeholders.
- There is need for examiners and item developers to go through rigorous training in order to maintain set standards.
- The education and examinations systems need constant monitoring and reviewing to ensure that the processes do not provide unfair advantages to some learners.
- Standards setters have a responsibility to ensure that assessment and moderation systems are realistic, workable, and practical for assessors and moderators.
- Syllabus developers should include ethno-based assessment strategies across all subjects to encourage use of local resources during teaching and learning processes.

References

- Black, P. and Wiliam, D. (1998) Inside the Black Box: Raising Standards Through Classroom Assessment. In *Phi Delta Kappa International.*
- Creswell, M.J. (1996) 'Defining, Setting and Maintaining Standards in Curriculum-Embedded Examinations: Judgements and Statistical Approaches'. In H. Goldstein and T. Lewis (Eds.),

Assessment-Problems, Development and Statistical Issues: A Volume of Expert Contributions. (p57-84). Chichester: John Wiley and Sons.

- Ewell, P.T. (2002) 'Grading Student Learning: You Have to Start Somewhere'. Retrieved 9 March, 2006 from http://measuringuphighereducation.org/2002/articles/peterwell.htm
- Gipps, C. (1990) 'The Debate Over Standards and the Uses of Testing'. In British Journal of Education. 36 (p 21-36).
- Goldstein, H. and Heath, D. (2000) *Educational Standards*. London: Macmillan.
- Harman, G. (2002) *Quality Assurance Initiatives in Thailand.* Armidale: International Higher Education.
- Jones, R.L. and Bray, E. (1986) *Guide to Assessment in Education*. In Assessment: From Principles to Action. Houndmills: Macmillan Education Ltd.
- Sanders, W.L. and Horn, S.P. (1995) 'Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes'. In *Education Policy Analysis Archives*, 3 (6), March 3, 1995.
- Secretary's Commission on Achieving Necessary Skills (SCANS) 1991 'What Work Requires of Schools'. Washington: Department of Labour.
- Siyakwazi, P.D. (1997) 'Reconceptualizationof the Home Economics Curriculum in Zimbabwe.' In the *Zimbabwe Bulletin on Teacher Education*, 5 (1), March 1997.
- UK NARIC (2002) Assessment of the New ZIMSEC 'A' Level Syllabi. Commercial in Confidence.

Assessing the proficiency of adult learners in reading literacy in South Africa

Surette van Staden and Sarah Howie

Abstract

This paper focuses on one of the South African national options for the Progress in International Reading Literacy Study (PIRLS), namely the assessment of adult literacy in private and public training centres in South Africa. As a developing country, South Africa has limited access to skilled workers and the tendency is for the majority of learners to drop out of formal education for various reasons. In order to alleviate poverty and to improve the social and economic development of the country, it is essential for policy makers to determine that learning acquired through formal schooling and adult basic education is comparable. In terms of achieving education for all in South Africa (a concept introduced at the Jomtien World Conference on Education for All in 1990), frameworks such as the NQF provide a sound basis for ensuring that learning outcomes are achieved through both sectors of education.

The ABET study would aim to determine the performance of ABET learners on reading literacy as well as establishing the extent to which performance in reading literacy among ABET learners are equivalent to that of Grade 4 learners. This paper presents an overview of the study's main design issues as well as some preliminary findings from the pre-pilot study.

Introduction

The aim of this paper is to analyze the extent to which assessment instruments designed and implemented at the grade 4 level in primary schools may be implemented for adult learners at the equivalent ABET level (Level 2) in order to monitor and evaluate adults reading literacy.

The importance of literacy has become evident in the 50 years since the United Nations declared it to be a basic human right along with the right to adequate food, health care, and

housing. Literacy education has indeed become a tool to help address what might be perceived to be more pressing needs for food, health care and housing.

The United Nations Literacy Decade was declared for 2003 to 2012. According to UNESCO statistics, about 861 million people (or about 20% of the world's adults) cannot read or write or participate fully or optimally in the organization and activities of their societies. Of these illiterate adults, 70% live in Sub-Saharan Africa, Southern and Western Asia, Arab countries and North Africa, while two thirds are estimated to be women .

For both industrialized and developing countries literacy education is near the top of the policy agenda. UNESCO estimated that by the year 2000 approximately 1 billion illiterates remain worldwide, while the prospects of a radical improvement within the foreseeable future seems to be unlikely. The concern is that as certain parts of the world have very high illiteracy rates, these are often the same countries neither without the means to implement wide-scale and effective literacy programmes nor to evaluate such programmes. However as literacy is of central importance to development , it is these countries for whom it is critical that solutions to the problem of illiteracy be found as increasingly, it is correlated with higher levels of income and job productivity.

Why assess adult learner progress?

In his foreword for the Education For All (EFA) Global Monitoring Report 2005, the Director General of UNESCO, Koichiro Matsuura mentioned that "the quest to achieve Education For All (EFA) is fundamentally about assuring that children, youth and adults gain the knowledge and skills they need to better their lives and to play a role in building more peaceful and equitable societies" (Matsuura, 2004: Foreword).

Following the Jomtien *World Conference on Education For All* in 1990, member states resolved to provide Education For All their citizens irrespective of age. Six goals and targets were set with intentions for them to be met by the year 2000. For adult basic education the first goal was to reduce the adult illiteracy rate (the appropriate age group to be determined in each country) by one-half its 1990 level by the year 2000, with sufficient emphasis on female literacy to significantly reduce the current disparity between male and female illiteracy rates. The second goal on adult basic education was to expand provision of basic education and training in other essential skills required by youth and adults, with programme effectiveness assessed in terms of behavioral change and impact on health, employment and productivity (Global Synthesis, Education For All 2000 Assessment, Unesco, p.13). Although these targets were set to be achieved by the year 2000, when the time came in 2000 at the Dakar conference they were reset with a deadline for 2015. The Dakar targets for adult basic education were to ensure that learning needs of all young people and adults are met through equitable access to appropriate learning and life skills programmes; and achieving 50% improvement in levels of adult literacy by 2015, especially for women and equitable access to basic and continuing education for all adults (EFA Global Monitoring Report, 2002: 13).

In South Africa, assessment projects such as the Third International Mathematics and Science Study (now Trends in International Mathematics and Science Study) TIMSS, Progress in International Reading Literacy Study (PIRLS) and Southern African Consortium

for Monitoring Educational Quality (SACMEQ) have been implemented to evaluate the quality of education provision as well as progress made in attaining specific educational goals. On the one hand they serve to benchmark the standard of education in this country with international standards. On the other hand they assist to develop education indicators in order to measure the extent to which policy goals are being met in practice to ensure accountability.

According to Crouch (a.n) education indicators are a recent phenomenon in South Africa. He argues that previously, education indicators were not used for the following reasons: firstly South Africa did not have a public vocation to accountability; secondly education was previously seen as a service rendered based on trust rather than as a co-investment to be co-monitored and co-managed via partnerships; thirdly the purposes of education were too implicit, opaque, differentiated, and contested to warrant a unitary system of monitoring and evaluation; fourthly the focus of the previous education system indicators was not on delivering value-for-money; and finally there was a lack of ability to produce raw data to elaborate on the indicators of the previous education system (a.n:2-3).

In a democratic South Africa the need for education indicators has been emphasized. According to Crouch this is because there is an idea that in a democracy there must be control mechanisms in order to self-regulate and adapt; there is also a need for political accountability and there is a need to include bureaucratic control mechanisms in a democracy. Therefore the education indicators serve as part of a feedback loop that leads to correction. They aid in detection; they also assist in establishing a social control system that is based on a measure of accountability (Crouch a.n.: 4).

While Crouch goes on to indicate the types of education indicators that have been used in education policy documents and legislation, his focus remains largely on domestic aspects that drive or represent the need for education indicators such as matric results. The reality is that even a country like South Africa has exposed itself to further control mechanisms that emanate from outside as a move to 'benchmark' itself internationally and to attract foreign aid and investment. Globalization has enabled policy borrowing between the South African government and other countries. According to Knight and de Wit (1997 in Woodhouse 2001:1) globalization refers to the increasing flow of technology, finance, trade, knowledge, people, values and ideas across borders. In education this is evident in the adoption of transnational policies such as the EFA.

The purpose of extending the PIRLS 2006 study to incorporate adult learners as well would be to determine the level of reading literacy amongst those adults attending Adult Basic Education and Training in public adult centers.

Through the National Qualifications Framework (the NQF), South Africa has an education framework that intends to compare and equate learning that was acquired through formal and non-formal schooling. As a developing country, South Africa has limited access to skilled workers and the tendency is for the majority of learners to drop out of formal education for various reasons. In order to alleviate poverty and to improve the social and economic development of the country, it is essential for policy makers to determine that

learning acquired through formal schooling and adult basic education is comparable. In terms of achieving education for all (a concept introduced at the Jomtien World Conference on Education for All in 1990) in South Africa, frameworks such as the NQF provide a sound basis for ensuring that learning outcomes are achieved through both sectors of education – this implies the same quality of education, thereby maximizing access to education for all.

The progress in international reading literacy (PIRLS) in South Africa

PIRLS is an international comparative evaluation of reading literacy of Grade 4 (9 year-old) learners involving more than 40 countries. It is run under the auspices of the International Association for the Evaluation of Educational Achievement (the IEA). As an organization, the IEA undertakes international studies that benchmark performance of school-going children in mathematics, science, civic education, information, communication, technology and reading to name a few.

PIRLS 2006 is the second, after PIRLS 2001, in a five-year cycle of assessment that measures trends in children's reading literacy achievement and policy and practices related to literacy PIRLS aims to provide trends and international comparisons on:

- The reading achievement of Grade 4 learners.
- Learners' competencies in relation to goals and standards for reading education.
- The impact of the home environment and how parents foster reading literacy.
- The organization, time and reading materials for learning to read in schools.
- Curriculum and classroom approaches to reading instruction.

Currently, 46 countries are involved in this collaborative analysis of children's reading literacy and the factors that influence reading acquisition. Whilst there are 30 European countries participating, only three African countries namely Egypt, Morocco and South Africa are involved. Several developing or newly emerging countries are also participating such as Indonesia and in the Middle East and Western Asia Israel, Kuwait, Qatar and Iran. The Americas are not well represented with only the USA, two provinces in Canada and Nicaragua taking part. Asian participants include China, Hong Kong, Singapore and Chinese Taipei. New Zealand is the sole participant from Australasia.

PIRLS focuses on three aspects of reading literacy (Campbell, Kelly, Mullis, Martin and Sainsbury, 2001), namely:

- Processes of comprehension.
- Purposes for reading.
- Reading behaviours and attitudes.

Processes of comprehension refer to ways in which readers construct meaning from text. According to the PIRLS Assessment Framework and specifications for the 2006 study, readers construct meaning in different ways when faced with the task of reading. They focus

on and retrieve specific ideas, they make inferences, interpret and integrate information and ideas as well as evaluate and examine text features.

Four types of comprehension processes are used in the PIRLS assessment to develop the comprehension questions derived from reading passages that are presented to learners. A range of questions, each dealing with a particular process, enables learners to demonstrate their abilities and skills in constructing meaning from written text. PIRLS distinguishes between the following comprehension processes:

- Focusing on and retrieving explicitly stated information, where retrieving appropriate text from a reading passage not only means that the learners needs to understand what is stated explicitly in the text, but to also ascertain how that information is related to the information sought.
- Making straightforward inference, where the construction of meaning from text requires of readers to make inferences about ideas or information not stated explicitly in the text.
- Interpreting ideas and information, an interpretive process, where learners attempt to construct a more specific or complete understanding of the text by integrating personal knowledge and experience with meaning found in the text.
- Examining and evaluating content, language and textual elements, which allows for reflecting on textual elements, such as structure and language, in order for the learner to examine how meaning is presented.

The PIRLS assessment focuses on two purposes of reading, namely:

- Reading for literary experience
- Reading to acquire and use information

These two purposes for reading account for most of the reading done by young learners in and out of school. In literary reading, the reader engages with the text to become involved in imagined events, settings, actions, consequences, characters, atmosphere, feelings and ideas. The main form of literary texts when reading for literary experience in PIRLS assessments is narrative fiction. When reading to acquire and use information, the learner does not engage in imagined worlds, but with aspects of the real world by means of informational texts. By means of informational texts, the learner can understand how the world is and has been and why things work the way they do. These passages are not only aimed at the acquisition of knowledge and information, but also to assess the learner's ability to use reasoning.

Although the PIRLS assessment distinguishes between these two purposes for reading, the underlying processes and strategies readers use for both are very similar. Each of these purposes for reading is often associated with certain types of texts. For example, reading for literary experience is often associated with fictional material, while reading to acquire and use information is more likely to be associated with informative articles and instructional texts. Although the PIRLS assessment takes the form of fictional passages when reading for the purposes of literary experience and articles for the purposes of reading to acquire and use information, the purposes for reading do not align strictly with these types of texts. Because tastes and preferences vary so widely, almost any text could meet either purpose.

Reading behaviours and attitudes refer to those behaviours and attitudes that would promote lifelong reading habits. The PIRLS assessment makes use of contextual questionnaires that are given to learners, Grade 4 teachers, principals and parents to gauge these attitudes and behaviours.

Following the field test that was done in March 2005, data collection for the main study took place during October and November 2005 when data was collected using a representative sample at approximately 400 schools throughout South Africa. Data was collected not only in terms of assessing learner performance on a number of reading literacy passages, but also collecting data by means of contextual questionnaires that were given to learners, teachers, school principals and parents.

The IEA encourages countries to expand their national study with so-called national options in which countries may administer a number of instruments or expand the international instruments with a number of questions addressing a number of issues or research questions that are important for the country.

In South Africa several additional components have been included in the South African PIRLS project to enhance the relevance of the project and address a number of aspects that are unique to our educational environment. These do not form part of the international part of the study.

A decision was made to include three such national options, namely:

1. Assessing learners for English Second Language proficiency.

Learners in South Africa begin their schooling through the medium of their mother tongue (one of the official 11 languages) and the medium only changes to either English or Afrikaans at the beginning of grade 4.The transition in the language of learning is seen as the root cause for many learners struggling to access information within the classroom. For the majority of learners (more than 70%), this additional language (usually English) is experienced as a foreign language as they seldom speak or hear the language outside the classroom. This is especially true in rural areas.

Therefore, learners were not only tested in their home language, but were also tested by means of an English Second Language test. This instrument was developed in collaboration with Grade 4 teachers and aims to assess learners' proficiency in reading English as an additional language.

2. Including Grade 5 learners for assessment.

The rationale for including grade 5 learners is to study the progression in reading ability from grade 4 to grade 5, given the transition of learners in the languages of learning in grade 4. The South African team tested grade 4 and 5 learners in all 11 official languages.

3. Assessing ABET learners using the PIRLS instruments.

The purpose of extending the PIRLS 2006 study to incorporate adult learners would be to determine the level of reading literacy amongst those adults attending Adult Basic Education and Training in public learning adult centers. The remainder of this paper will focus on an

overview of the study's intended use of the PIRLS instruments to assess adult learners as well as some preliminary findings from the pre-pilot study.

Using PIRLS 2006 instruments to assess adult learners

The first phase of the ABET project aimed to evaluate the extent to which the assessment instruments administered to Grade 4 learners could be used to evaluate the performance of adult learners at the equivalent level. This research used selected reading passages, as used in the PIRLS 2006 study on Grade 4 learners to assess ABET learners' performance in reading literacy.

Essentially, the ABET study aimed to determine the following:

1. What is the performance of ABET learners on reading literacy?
2. What is the role of the educator in facilitating the acquisition of reading literacy in ABET learners?
3. To what extent is ABET learner performance on reading literacy attributable to eco-systemic factors?
4. To what extent is performance in reading literacy among ABET learners equivalent to that of Grade 4 learners?

With regard to the fourth question, in South Africa it has been assumed that the general education and ABET programmes are comparable and that a qualification obtained under ABET is equivalent to the general education band. Empirical research has not verified this. While the PIRLS 2006 project is mainly focused on comparing our Grade 4 reading literacy to those of Grade 4 learners in other countries, it has not yet been determined that the curriculum and learning outcomes for Grade 4 are purported to be equivalent to those of ABET Level 2 and are in fact comparable.

The objectives for the study can be outlined as follows:

- To conduct a baseline survey, assessing adult learners in order to determine levels of reading literacy. Apart from ascertaining levels of reading performance, contextual data will also be collected in the form of administering questionnaires to learners, lecturers and heads of ABET centers.
- To analyze collected data, aimed at, amongst others, making comparisons between adult learner performance and Grade 4 learner performance.
- To write a comprehensive report outlining the findings and identifying possibilities for further research.

As stated in earlier sections, reading for literary experience involves the use of fictional passages, each with a number of questions. For the purposes of the ABET assessment, two passages aimed at reading for literary experience were deemed to be more appropriate for use with adult learners. Both passages tell fictional stories, but with a moral underpinning and the content of the stories is such that it would appeal to a more mature audience, not only to children. Following each passage, a number of multiple-choice and constructed response questions are asked. Questions each deal with a particular comprehension process as described and explained in section 3.

Section 3 also describes how, by means of the informational passages, the learner does not engage in imagined worlds, but with aspects of the real world through informational texts. By means of informational texts, the learner can understand aspects of the world by means of factual texts. These passages are not only aimed at the acquisition of knowledge and information, but also at assessing the learner's ability to use reasoning. Two appropriate informational passages were selected for the purposes of assessing adult learners.

Experts in the field of adult education evaluated the selected passages in order to assure the appropriateness of their use on adult learners. Amongst others, the passages were evaluated as follows: In terms of the length of the passages, they were found to be more than double in the length of the reading passages that are in some Level 2 English Modules and twice as long as most of the reading passages in the IEB English Level 2 examinations.

The proposed passages were evaluated favourably and could be used for ABET learners. Although they are not directly related to their lives, the situations discussed in the informational passages might not be totally outside the experience of South African ABET learners. Learners should be able to use their existing knowledge to make sense of the passages—many South African adults will know something about sharks and shark attacks on swimmers. Likewise, the hiking passage could be of interest to some ABET learners, many of whom would be able to relate to the discussion on physical fitness.

The literary reading passages were evaluated as texts that could be understood as stories, but that also act as an allegory for human behaviour and a means of teaching children social and moral skills. The first selected passage was appropriately evaluated as an African folk tale that could appeal to adults, while the second passage could be problematic, since it is a children's story told from the perspective of the animals that can talk. The reference to wolves in the story might also not be a familiar animal to which learners in South Africa would be able to relate.

While simple sentences are used in all the passages, there are also many complex and fairly lengthy sentences occurring in the passages. For adult Level 2 readers, long complex sentences that contain too many ideas should be avoided.

Evaluators identified many words and phrases that would be unfamiliar to ABET Level 2 learners e.g. *pandemonium*, *perplexed*. ABET learners would not use them in their daily lives nor would they be likely to come across them when reading, listening to the radio or watching television. It was suggested (similar examples are found in IEB exams papers) that a glossary of unusual or difficult words and phrases is provided as this helps ABET learners to make sense of the text. Without such a learning aid, ABET learners often waste valuable time unsuccessfully trying to make sense of a particular word or phrase and thus may become very frustrated and demotivated. Level 2 learners and above do not always read through the whole passage and try to infer the meaning of unknown words and phrases from the rest of the text.

Based on the evaluation of these passages, an outline and detailed description of the types of comprehension processes referred to in the PIRLS assessment. This was used as a basis for devising a breakdown of ABET Unit Standards, Specific Outcomes, Assessment Criteria and specific levels at which these occur for each of the items in each of these passages. The assessment framework therefore not only makes reference to the PIRLS framework, but also links each item to the ABET Unit Standards and Assessment Criteria for the purposes of this study.

Results of pre-pilot of adult learners using PIRLS assessment

In light of predictions from various experts in the field and the above mentioned criticism on the appropriateness of the use of PIRLS instruments in the assessment of ABET learners, a pre pilot study was conducted early in March 2006. The rationale behind the pre pilot was not to gather data from a scientifically selected, representative sample, nor was it to use the results from gathered data to make generalizations. Instead, the pre pilot was merely conducted to gauge how a small sample of adult learners would react to and perform on the PIRLS instruments. This would provide preliminary 'evidence' that would either support or refute the predictions made by experts in the field.

The following section will aim to describe the procedures followed during the pre pilot, as well as some preliminary results.

Sample of learners assessed

Early in March, 2006 a sample of 30 ABET learners from a public Adult Education Centre situated in Katlehong on the East Rand of Gauteng were assessed in English using the selected PIRLS passages described previously. The centre serves a wide area of learners, both school going and from the ABET sector. The principal of the centre agreed to allow the assessment of the learners therefore the selection of this centre is based purely on convenience and not on scientific selection.

The adults participating in the assessment ranged in ages from as young as 14 to as old as 60 years of age. In total, 30 learners were assessed, 11 of who were male with the remaining 16 being female. Of these 30 learners, 12 are in the process of completing ABET Level 3 education—the other 18 learners are all on ABET Level 4. This sample of learners does not match the intention of assessing adult learners at ABET Level 2. Due to unforeseen circumstances, Level 2 learners were not available for testing on the evening of the visit to the adult learning centre. When Level 4 learners availed themselves for testing, a decision was made to test these learners instead of returning from the field empty handed. All these learners have employment as blue collar workers and their motivation behind attending ABET classes ranged from purely self-improvement reasons to being better able to procure more suitable employment.

Procedures followed during the assessment

The same procedures as prescribed by the IEA for data collection during the PIRLS assessment of Grade 4 learners were followed during the assessment of the adult learners. Each learner was given a test booklet, containing two reading passages with questions to complete. It was ensured that learners sitting next to each other received different test booklets—in this way, any possibility of learners copying from one another are eliminated.

Directions (as provided of the first page of each test booklet) were explained to learners orally. Particular attention was paid to example questions to ensure that learners understood the type of questions they would be expected to answer during the assessment. Strict time-keeping procedures followed the moment learners were ready to begin the assessment. In order to complete the first passage, 40 minutes was allowed. Learners were not allowed to continue with the second passage should they complete the questions to the first passage within the allocated time. After the first 40 minutes elapsed, learners were allowed a 10-minute break—although learners were allowed this break, nobody made use of the opportunity and testing continued for the second passage. Again, 40 minutes was allowed for the completion of the second reading passage. After the elapse of this second session, the assessment was concluded.

For the duration of the assessment, learners were allowed to ask questions should anything be unclear. Very little assistance however was allowed, especially where learners requested assistance with their answers. In most cases, learners could only be told to write down what they think would be right or appropriate.

Results from implementation of PIRLS instruments

The title of this paper may lead the reader to believe that this paper will provide a comprehensive account of adult learner performance. Rather, the emphasis is on the quality and appropriateness of the intended use of PIRLS instruments on adult learners and not adult performance as yet.

In light of the small sample size, results of adult performance on the selected reading passages will be restricted to a discussion of item difficulties. These difficulties should provide an indication of the proportion of adult learners who were able to provide correct answers to the questions they were asked on each of the passages.

Of the 30 learners who were assessed, not many were able to complete the questions given the allocated time in which to do so, therefore booklets were returned with a very high non-response rate. For the purposes of this pre-pilot it was decided to keep to the prescribed times as set by the PIRLS assessment (40 minutes for each passage). In addition, time was constrained since learners were only available from 18:00 to 20:00 in the evening.

Non-responses imply that learners simply ran out of time and couldn't respond to the questions, or learners simply did not make any attempt to answer the questions, even if they had enough time to do so. For this reason, difficulty values will be reported on in two ways—firstly, those values derived from all learners (responses and non-responses) and secondly those difficulty values as derived from only those learners who attempted to provide answers (non-responses therefore excluded from the calculation). Items with a significant improvement in difficulty value (for example in cases where the item changes from a difficult to an easy item) when calculated in these two ways are highlighted in the tables below.

Results for Booklet 1 containing 'Reading for literary experience' passages

A total of 15 learners answered questions contained in booklet 1. These learners ranged in age from 18 to 53 and included 5 males and 7 females. The first reading passage consisted of 7 multiple-choice items and 5 constructed response items. The second passage also consisted of 7 multiple-choice items, but 6 constructed response items.

Table 5: Results for Literary Passage 1

Item Number:	Type of Item:	N	Difficulty Value (including non-responses):	N	Difficulty Value (non-responses left out):
1	MCQ	15	0.33	5	0.90
2	MCQ	15	0.20	5	0.60
3	MCQ	15	0.06	5	0.80
4	MCQ	15	0.13	6	0.33
5	CR	15	0.10	4	0.37
6	MCQ	15	0.06	5	0.20
7	CR	15	0	3	0
8	MCQ	15	0.06	6	0.16
9	CR	15	0	3	0
10	CR	15	0.13	3	0.66
11	MCQ	15	0.06	4	0.25
12	CR	15	0.06	2	0.50

- MCQ – Multiple Choice Question
- CR – Constructed Response

Table 6: Results for Literary Passage 2

Item Number:	Type of Item:	N	Difficulty Value (including non-responses):	N	Difficulty Value (non-responses left out):
1	MCQ	15	0.13	12	0.16
2	MCQ	15	0.20	12	0.25
3	CR	15	0	11	0
4	MCQ	15	0.20	10	0.30
5	MCQ	15	0.13	11	0.18
6	MCQ	15	0.40	11	0.54
7	MCQ	15	0.26	11	0.36
8	MCQ	15	0	11	0
9	CR	15	0.16	8	0.31
10	CR	15	0.06	10	0.10
11	MCQ	15	0.13	9	0.22
12	CR	15	0	11	0
13	CR	15	0.11	7	0.35
14	CR	15	0.10	9	0.16

- MCQ – Multiple Choice Question
- CR – Constructed Response

Ideally, item difficulty values should include items from a range of easy, medium to difficult items. Easy items would typically be those that more than 70% of learners are able to answer, while medium items would include those items where between 40 and 60% of learners are able to provide correct answers. Difficult items fall in the category of 30% and lower and would therefore include those items that only a few learners are able to answer correctly.

Tables 1 and 2 indicate difficulty values that would suggest very high difficulty values for each of the items found in the two reading passages contained in booklet 1. This means that a very low proportion of learners were able to answer any of the given items correctly. Values range from 0 (where none of the learners were able to provide the correct answer to a specific item) to only as high as 0.40 (implying that only 40% of learners were able to answer an item correctly). A slight improvement in difficulty values occurs when those learners who failed to respond are taken out of the equation. An increase in difficulty value in these cases, however, should be interpreted with caution. Learners who actually responded to items were very few. When calculating difficulty values when only making use of those learners who responded, a difficulty value is derived from literally only 3 or 4 responses. Therefore, the marked increase in performance on some items should not be interpreted in isolation from this fact.

Results for Booklet 2 containing 'Reading to acquire and use information' passages

As was the case with booklet 1, a total of 15 adult learners were assessed using booklet 2, which contained two informational passages. Learners' ages ranged from 14 to 60 years of age, 6 of who were male and 9 of whom were female. Also consisting of a number of constructed response items and multiple-choice questions, results for these two passages are as follows:

Table 7: Results for Informational Passage1

Item Number:	Type of Item:	N	Difficulty Value (including non-responses):	N	Difficulty Value (non-responses left out):
1	MCQ	15	0.13	10	0.20
2	CR	15	0.46	13	0.53
3	CR	15	0.26	13	0.30
4	MCQ	15	0.60	13	0.69
5	MCQ	15	0.46	13	0.53
6	MCQ	15	0.26	13	0.30
7	MCQ	15	0.20	13	0.23
8	CR	15	0.40	11	0.54
9	MCQ	15	0.33	10	0.50
10	MCQ	15	0.33	10	0.50
11	CR	15	0.36	9	0.61
12	CR	15	0.16	8	0.31

- MCQ – Multiple Choice Question
- CR – Constructed Response

Table 8: Results for Informational Passage 2

Item Number:	Type of Item:	N	Difficulty Value (including non-responses):	N	Difficulty Value (non-responses left out):
1	CR	15	0.40	7	0.85
2	CR	15	0.20	6	0.50
3	MCQ	15	0.06	6	0.16
4	MCQ	15	0.13	5	0.40
5	CR	15	0.26	4	0.90
6	MCQ	15	0.20	5	0.60
7	CR	15	0.16	4	0.62
8	MCQ	15	0.06	5	0.20
9	MCQ	15	0.06	5	0.20
10	CR	15	0.02	3	0.11
11	MCQ	15	0.06	4	0.25
12	CR	15	0.06	4	0.25

- MCQ – Multiple Choice Question
- CR – Constructed Response

Similar to the passages in booklet 1, the performance on items in passages in booklet 2 was very poor. With the exception of item 4 of informational passage 1 (where 60% of learners provided the right answer), item difficulties again are well below 40% in most cases, meaning that very few learners were able to provide correct answers to any of the given items. With the removal of non-responses, item difficulties seem to improve, thereby implying that learners who provided responses to the questions were able to either answer the multiple choice questions correctly, or were able to obtain full or partial credit for the constructed response items. As mentioned earlier, this improvement should be interpreted with caution given the low number of responses the difficulty value is derived from.

Discussion

The aim of this paper was to analyze the extent to which assessment instruments designed and implemented at the grade 4 level in primary schools may be implemented for adult learners at the equivalent ABET level (Level 2) in order to monitor and evaluate adults reading literacy.

As mentioned in earlier sections, this pre pilot study was merely conducted to gauge how a small sample of adult learners would react to and perform on the PIRLS instruments. Given the small sample of adult learners that were tested, the above mentioned results of their performance on the selected reading passages were restricted to a discussion of item difficulties. Although basic in nature, these item difficulties provide a preliminary indication of how adults might perform on the PIRLS instruments.

When looking at the information provided in table 1 to 4, it becomes clear that the adults in this small sample found the items to the passages very difficult. Very few learners were able to provide correct answers on any of the given items, regardless of whether items were multiple choice type questions or constructed response items. A high non-response rate also applies to this sample, where many booklets returned without the learners having even attempted to answer the questions. In fact, of the 30 learners, only 2 learners managed to complete the questions to both reading passages.

Some observations were made during and after the assessment that deserve mention here. Firstly, adult learners seemed to ask the same type of questions than those asked by the grade 4 learners. For example, during data collection for the main study, grade 4 learners often put up their hands and asked assistance in answering the multiple choice questions. So too did they often inquire as to whether they were supposed to answer the questions after reading the passage. Another frequent occurrence was found where learners wrote 'answers' in the spaces provided in the example questions – even when emphasizing that those only served as examples, some learners still completed these questions. These are but few examples of what frequently occurred among the grade 4 learners. These types of questions however also occurred in exactly the same fashion among the adult learners in this sample. Of note was also the fact that adult learners (to a greater extent than the grade 4 learners) didn't seem to understand that the questions asked after each passage pertained to that reading passage. Purely based on observations made during the assessment it became clear that the adult learners did not seem to be familiar with comprehension tests, despite assurance given by their teacher that these learners have in fact been exposed to it. It also seemed that adult learners were less comfortable to work on their own and less familiar with completing a task independently of others.

Upon completion of the assessment, learners were asked about their experiences of the test. Interestingly, the majority of learners indicated that they found the test rather easy and enjoyable and most of them expected to do well. There seems to be somewhat of a gap in terms of how learners perceive their own performance to what it actually is. Learners were also asked whether they would have preferred the test to be in their mother tongue (in this case some Zulu and Tsonga learners were present). Unanimously, learners indicated their preference for English, saying that it is the language of business, which they should get to know.

In light of the results that seem to indicate that adult learners find the PIRLS passages very difficult to complete, the way forward for this project lies in designing an assessment instrument where aspects of basic and functional literacy are included as well. In assessing basic and functional aspects of literacy, one might be able to gain more insights in terms of gauging where adult learners are and how well they can read. The results of such an instrument might prove to be more meaningful instead of exposing adult learners only to the PIRLS passages, which they do not seem able to do.

For the remainder of the year, a pilot project using a larger sample of adult learners is planned, using an instrument that would incorporate both items to measure functional and

basic literacy as well as selected PIRLS passages. A pilot study would allow for the refinement of the instrument, upon which a full scale, main study could be conducted within 2007.

Diffusion of the National Qualifications Framework and outcomes-based education in Southern and Eastern Africa

Linda Chisholm, Human Science Research Council

Introduction

One of the major debates in comparative education at present is whether there is increasing convergence or divergence in national systems of education. In the institutionalist perspective, homogenisation and 'institutional isomorphism' is striking around the world. This sameness is most remarkable 'on two fronts' (Meyer & Ramirez, 2003, p. 127): that of credentials and standards on the one hand and that of the educational definition of the world on the other. The counter argument sees not homogenisation and standardisation, but indigenisation and modification through a much more complex process of the 'displacement' of reforms by local realities (Steiner-Khamsi, 2003). In this process the 'semantics of reform' (Schriewer, 2000a) provide the appearance of similarity. Linked to this approach is the view that developing countries are often 'laboratories of modernity' and 'test sites for contested school reforms' (2004, p. 183; see also Stoler, 2002). Increasingly, this and other comparative approaches have been urging a more historical approach to analysis of educational borrowing (Ochs and Phillips, 2004; Crossley and Watson, 2003; Watson, 2001) and have also begun to emphasise the nature and character of regional borrowing (Dale and Robertson, 2002).

These approaches are particularly important for understanding the nature of education policy and curriculum diffusion in southern and eastern Africa where countries are both locked into historical relationships of varying kinds with one another and with others beyond the African continent. They are thus open to diverse influences from within, around and without. This article takes the case of the National Qualifications Framework and outcomes-based education to explore the diffusion of educational ideas in regional contexts further. The centrepiece of education and curriculum policy in South Africa when it democratised in 1994, the National Qualifications Framework and outcomes-based education, have diffused throughout the region. Both Jansen (2004) and Spreen (2004) have provided accounts of

how outcomes-based education became adopted and indigenised in the South African context. There has been less research into the nature of corresponding changes in the southern and eastern African region as a whole (Chisholm, 2005). There is also little understanding of whether ideas have spread because of, despite or in tandem with one another, in a process of mutual attraction (Ochs and Phillips, 2004) or whether the influence of external agencies in the developing world context is more telling in ensuring their spread (Samoff, 1999).

The work of Jansen (2004) and Spreen (2004) reveal an active local process of policy appropriation through individuals networked with one another through history, organisations and international linkages. In this process, rationales for the adoption of particular policies, decontextualised of their own origins and histories, are developed and policies are rearticulated to new needs and desires. Schriewer has used the concept of 'discourse coalitions' in attempting to understand ways in which 'alliances are established between certain representatives (or groups of representatives) of the social sciences, on the one hand, and certain political actors (or groups of political actors) on the other (Schriewer, 2003, p. 74). The mutually reciprocal relationship results in the elaboration of new discourses. In this paper, the concept is broadened to include political and social actors who include not only social scientists but also intellectuals who may not necessarily be university based, but are based in non-governmental organisations.

Theoretically and methodologically, the concept of 'discourse coalitions' and the techniques of network analysis provide a methodology for understanding processes of lending and borrowing. The diffusion of ideas is also critically linked to the relationships of individuals to particular networks. Ideas spread through those with weak rather than strong ties, those on the peripheries interacting with other networks (Granovetter, 1983). The role of policy entrepreneurs in this process of diffusion is also pivotal, linking with the notion of policy diffusion as export.

This article argues that as far as the National Qualifications Framework is concerned, South Africa has constituted an experimental test site, a laboratory for new ideas. The idea was first developed in the context of training markets becoming a central policy objective in the strongly market-oriented societies of the UK and Australia (Atchoarena and Delluc, 2002, p. 34). The reasons for adopting it in the region have been different from the original reasons for South Africa's adoption of the National Qualifications Framework. The paper will argue that these ideas are spreading in the region at the very moment that South Africa appears to be taking a step backwards from them and questioning their efficacy. Although there has been a common 'semantics' of reform, however, there has been relatively little implementation and corresponding impact on practice.

But in order to understand regional processes of borrowing and lending in relation to the National Qualifications Framework and outcomes-based education, it is necessary first to outline briefly what the National Qualifications Framework and outcomes-based education are. The paper will then provide a brief history of the trajectory of educational change in the region to contextualise the discussion of borrowing and lending in the post-1994 period. An outline of South Africa's current political and economic role on the continent will follow,

and precede an exploration of some of the processes of borrowing and lending related to the National Qualifications Framework and outcomes-based education in the region. Using evidence from regional conferences, the article will argue first that conferences are particularly important sites for discourse coalitions and policy entrepreneurs to spread ideas, but they do not necessarily however have an impact on what actually changes in practice. They are an important but not the only source of new ideas. In so far as the author of this article is based and working in South Africa, it participates in this borrowing process. In the development of its own approach, this article consciously borrows international ideas and uses international references in order to develop the argument.

The National Qualifications Framework was introduced through the South African Qualifications Authority Act in 1995 as a means of addressing inequalities in learning across different racial groups in the society. The Act provides for an outcomes-based system in which there is an explicit focus on what has been learned, as measured against 'socially agreed standards.' A National Qualifications Framework and outcomes-based education are implicated with one another and assume adherence to a broader market-oriented policy agenda (Acthoarena and Delluc, 2002, p. 318). The National Qualifications Framework in South Africa formed part of a larger effort to address unemployment through regulating the labour market and process of skills acquisition through transforming the fiscal, institutional and legislative framework within which institutions operate. In practice, as its chief architect Adrienne Bird, trade unionist and later Labour Department official, wrote in 2000, it:

Provides for the registration of general education as well as occupationally oriented education and training standards at general, further and higher levels of learning. It is made up of eight agreed national levels and a range of learning progression routes. Agreement on standards to be registered is recommended by 12 National Standards Bodies made up of employers, trades unionists, government officials, professional bodies and education and training providers…. Once standards and qualifications are registered, the quality assurance of the standards is ensured by a second set of established and new institutions. They register assessors and keep a record of learner achievements. They also formally accredit providers as competent to train…. The model was proposed to enable both young and older learners to have their current learning recognized through a process of Recognition of Prior Learning, and then have the opportunity to progress further in learning, be it in full-time or part-time contexts. (p. 166)

The National Qualifications Framework arose out of a transitional, pre-1994 context and desire to address unemployment, economic growth and learning. It was a consensus-building process between the trade unions, business and the African National Congress focused in the first instance on upgrading and improving skills of the unemployed, workers and older people. But in an ambitious early move, it assimilated into its sphere the entire schooling and higher education institutional complex. This was a process attended by considerable conflict, as will be shown. It was within the South African context, within the region and arguably internationally, a major educational experiment that ultimately had more impact on the discourse than the practice of education and training; policy intentions, with some minor exceptions, did not easily translate into implementation, whether in the school, higher education or training arenas (see, for example, Atchoarena and Delluc, 2002, p. 340).

Educational change and development in the SADC region in historical perspective

Educational change and development in the region has been marked by different periods of history: the period of colonialism and struggle for independence, the immediate post-independence period, structural adjustment, and the political and economic liberalisation of the post-1990 period. These periods do not coincide exactly with one another, but each phase has had distinctive impacts on the goals and directions that education has taken (The information in the following paragraphs rely on Chisholm, Makwati, Marope, Dumba-Safuli, 1998). Thus the colonial period in southern and eastern African countries saw the development of education as an elite phenomenon. Independence was achieved at different times in the region: Botswana, Lesotho, Swaziland, Zambia, Tanzania, Malawi and Mauritius in the early 1960s, Angola and Mozambique in 1975, Zimbabwe in 1980, Namibia in 1990 and South Africa in 1994. Modernisation and expansion of education at all levels in order to achieve higher, national and individual emancipatory goals accompanied the achievement of independence. But for those who achieved independence in the 1960s and 1970s, the period 1975-1985 was marked by a protracted political and economic crisis both in the region and in the world economy. As terms of trade for primary products declined, so foreign exchange shortages emerged, reliance on foreign loans increased and indebtedness grew. Except for Botswana, no country in the region escaped this growing poverty trap.

But independence in the former Portuguese colonies did change the balance of power in the region against the white minorities in the then-Rhodesia, South West Africa and South Africa. Combined with poor management of resources, shrinking national budgets resulted in a squeeze on resources for education. Many countries in the region, except for Botswana – the one exception to the pattern - adopted structural adjustment policies from the mid-1980s in an attempt to address worsening conditions. These policies aimed at achieving macro-economic stabilisation but in fact the restraints on public expenditure, privatisation and decentralisation of services did not lead to improvements in living conditions and social services such as education.

Difficulties of the previous decade appear to have been compounded in this decade. Across the region, there was a drop in productive investments and there were few new enterprises. Greatest growth was recorded in the informal sector (Vieira, Martin and Wallerstein, 1992). The trend towards falling educational expenditure and greater reliance on donors grew. In some countries, external donors also began to finance recurrent costs, taking over basic state functions such as payment for staff salaries, equipment, furnishings, transport and security services. At the same time, countries in the region offered a home to South African and Namibian resistance movements, thus increasing their vulnerability to attack by South Africa.

A changing global order in the early 1990s, linked to a combination of local internal political dynamics and regional shifts as Namibia and South Africa gained independence, resulted in a process of political and educational reform across the region. Between 1992 and 1995, all countries in the region held multi-party elections. In each case, almost immediately following

independence, a process of educational reform began. In each case, the process of stakeholder involvement and discussion was central in legitimating the process. The Education for All conference held in Jomtien in 1990 provided the frame for most of these new reform initiatives: across the region, whether in newly-independent South Africa, or in Zambia or Tanzania, policy priorities became focused on improving educational access, quality, equity, efficiency and relevance. Although not a major focus of educational reform or the EFA agenda, learner-centredness (distinct from outcomes-based education but sometimes associated with it) also now explicitly entered the picture. The conference brought together donor agencies, national ministries of education, NGOs and researchers in support of realising the EFA goals within a framework of an economic order in which market principles held increasing weight: privatisation, cost-sharing and fiscal discipline were the financial counterparts to the educational reforms.

A 1998 SADC study of these policy formulation and implementation processes (Chisholm, Makwati, Marope, Dumba-Safuli, 1998) concluded that there were major obstacles to implementation, mainly at the level of physical and human resources and capacity. It recommended an annual forum be established to promote regional cooperation and collaboration between different role players in order to begin to address these deficiencies. The report suggested that despite the good intentions of policy, the EFA policies remained at the discursive level, as a 'semantics of globalisation', (Schriewer, 2000) a way of talking about education, but having little impact on it. A recent World Bank report on curricula, examinations and assessment in sub-Saharan Secondary Education (CASASE) (2005) provided evidence that although learner-centred education is increasingly promoted as part of this new educational reform movement (eg Botswana, Tanzania, Namibia, South Africa, Uganda), there is little sign of it in the classroom (Leyendecker, 2005, pp. 26/7).

However, the very exercise of an investigation into SADC education policy, planning and management initiatives in 1998 involved the bringing together of new and different networks in the region, whose key elements begin to tell the story of diffusion. Each person in the team was part of a powerful network in his or her own context. With weak ties to one another, and interacting with other networks, the exercise was an example of how policies are diffused through relatively isolated networks. In this instance, the donor (Netherlands Embassy) brought together an international (UNESCO) and regional agency (SADC) and researchers from four different countries who straddled government and research organisations.

Up to this point, the article has focused on how the discourses of policy had, by 1994, converged across the region as consequence of a combination of economic, political and social forces. It is now necessary to sketch South Africa's role in the region before and after 1994.

South Africa's current role on the continent

Under apartheid, the regional and foreign policy role played by South Africa was an aggressive and imperialistic one. Since the advent of democracy, there is some debate as to whether it is playing the role of regional hegemon or not. Supporting this perspective is the increasing economic penetration of Africa and especially the SADC region by South African

companies and parastatals. According to Daniel, Naidoo and Naidu (2003), South Africa is now the largest foreign direct investor in SADC, its foreign direct investment in the region exceeding that of the United States and the United Kingdom combined. In 2001, total South African foreign direct investment amounted to R26.8b (Daniel, Lutchman and Naidu, 2005, p. 549). Mozambique represents the fastest-growing area of South African economic involvement, seeming 'poised to eclipse Zimbabwe as South Africa's largest trading partner.' While South Africa's export trade to Africa has flourished, the same cannot be said of its African import trade – the trade balance in South Africa's favour is substantial (pp. 546/7). South Africa's major trading partners are Zimbabwe, Mozambique, Nigeria, Zambia and Angola. The most significant involvements are represented by the mining sector, banking and financial services, the telecommunication sector, the aviation market and Shoprite Checkers (553/4).

Despite this growing economic role in Africa, Daniel, Lutchman and Naidu do not believe that South Africa can be described as a regional hegemon. They distinguish between the Mandela and Mbeki Presidencies – in the former, foreign policy was often contradictory and still tied to past practices, as exemplified in the Lesotho incursion; in the latter, South Africa's involvement has been shaped through the African Union (AU) and New Partnership for African Development (NEPAD) and a policy framework that emphasises negotiation, partnership and cooperation. Not one of the critics of South Africa's 'quiet diplomacy' stance on Zimbabwe, for example, has called for war (Sachikonye, 2005). South Africa arguably does not exercise political and ideological hegemony in the region, although its role is substantial. In this context, it is to be expected that there would also be considerable exchange on questions of education and education policy.

Educational borrowing and lending in the region: the case of the National Qualifications Framework (NQF).

Not surprisingly, the first Southern African Development Community (SADC) Protocol on Education and Training adopted in 1997 promoted one of the key new features of the South African education and curriculum landscape: an integrated qualifications framework. Far from being just a technical attempt at coordinating qualifications, the National Qualifications Framework embeds an entirely new approach to education known as outcomes-based education. Outcomes-based education and standards-led education reform movements are in some contexts seen as distinct. The specifically South African variant is a version of a standards-based approach. South African trade unionists borrowed the idea from Canada, England, Scotland and New Zealand where national qualifications frameworks were nascent in the 1980s. The idea appealed as it promised a number of benefits, amongst them improving access to education through overcoming divisions between mental and manual labour, integrating education and training, recognising prior knowledge, and establishing equivalence between different qualifications to enable rapid upward mobility through different educational institutions and into the world of work. Although there is policy convergence in SADC on this question, however, the motivations for adoption differ from those of South Africa: 'For SADC, the development of compatible national qualifications frameworks represents a strategic instrument to increase the competitiveness of the sub-region and contribute to further economic and labour market integration' (Atchoarena and Delluc, 2002, p. 341.)

In 1997, the South African Qualifications Authority (SAQA) placed outcomes at the centre of new curriculum and assessment development in education and industry training. The interpretation and enactment of outcomes-based education has unfolded in different ways in the different sectors. Industry training took on competency-based unit standards in its curriculum model. The schooling sector rejected this and adopted a modified outcomes-based whole qualification approach. Higher education spurned outcomes and performance-based education in their entirety. This process has been attended by considerable contestation over the meaning and purposes of education.

The challenge and diversification of response unfolded through a series of steps. Only three years after it was introduced, in the face of seemingly intractable implementation problems, Curriculum 2005, as it was called, was reviewed (DOE, 2000). The Review found the approach to curriculum made little difference to what actually happened in the classroom, and well-resourced schools better able to adopt learner-centred education and new assessment methods than poorer-resourced schools. All schools complained that an excess of paperwork accompanied the new curriculum. The Review was a thinly disguised attack on outcomes-based education, but for political reasons the discourse remained intact.

In the face of considerable hostility from both the schooling sector as well as higher education, the National Qualifications Framework was reviewed shortly thereafter. (DOE, 2002) This Review in effect sought to reverse some of the more ambitious attempts at restructuring that were entailed in the panoply of standards-setting structures and bodies to which the South African Qualifications Authority had given rise. Although an outcomes-based, learner-centred approach still frames South African education policy and curriculum, some of its institutionalised features, such as the Standards-Generating Bodies established under the South African Qualifications Authority, have been dismantled. Unit-based qualifications for schools have also been abandoned. In school curricula, the 1990s emphasis on outcomes at the expense of subject knowledge and portfolio assessments instead of tests and examinations, have been replaced with a renewed emphasis on content and examinations (Vinjevold, 2005). Outcomes-based education in South African schools now appears to exist more in the breach than in the observance.

The South African experience of outcomes-based education has not been one that deserves export, and yet it is being exported. As in its own case, the reasons for import are local and specific, the sources of the idea diverse – including both South African and more international personages, agencies and institutions – and it is to be expected that the export will be modified, adapted and resisted along the way. The South African story of the NQF is very much one of the adoption of an idea that seemed to provide solutions to local problems, implementation in a context unable to absorb its full implications (Spreen, 2004). In a seminal work, Tyack and Cuban (1995) have shown how reforms introduced with great fanfare at one point can either take root, or wither away and die, forgottten, or be grafted onto existing institutions as add-ons. In the case of the NQF, it seems that there has been a process of selective appropriation, transformation and rejection of the original adopted and elaborated idea. As Allais has argued, the location of the NQF within a democratic discourse in South Africa has meant that its neo-liberal origins have been obscured, and that its basic

assumptions have been insulated from criticism even though it is not achieving its mission (Allais, 2003).

By 2002, a number of SADC countries were formulating qualifications frameworks with a strong vocational and training slant (Atchoarena, 2002). Other than South African and Namibia, Botswana, Mauritius, Zimbabwe, Malawi, Zambia and Mozambique had either developed a framework or established an authority or department to further develop and implement it by 2003 (SAQA, 2003, p. 18). These national qualification frameworks were being 'duplicated' 'with no factual evidence on implementing conditions, costs and on impact' (Atchoarena, 2002, p. 342.)

A Regional Qualifications Framework was also brought into being in the early 2000s (*Sowetan*, 2 September 2004; SAQA 2003). A SADC Technical Committee had been established and begun work. A feasibility study was being conducted with support from UNESCO and the ILO with the purpose of resulting in a project proposal for technical assistance in building national and regional capacity in building regional frameworks (SAQA, 2003, 20). This process paralleled the penetration of the region by publishers and nongovernmental organizations (NGOs) from South Africa. The latter 'appear to share the spirit of South African companies and entrepreneurs who are currently carving out export and investment opportunities throughout the continent' (Daniel et al., 2003; Morrow 2004, p. 332). According to Morrow, many of the most vigorous South African education NGOs are expanding 'into Africa'. (Morrow, 2004, p. 332). South African consultants are hired to advise on, amongst other things, curriculum reconstruction in the region for the purposes of developing the qualifications and curriculum frameworks and outcomes-based education (Morrow 2004; Pampallis, 2004). And publishers appear to be seeking new markets and opportunities for their products.

But what does all this mean in terms of the framework of borrowing and lending established above? I want to explore how and why the National Qualifications Framework has diffused into the region and with what effect. I want to argue that the experiment of the idea of the National Qualifications Framework has been exported, but there has been only partial acceptance. This export has happened at the moment of it being questioned in South Africa and where international support for it is diminishing. I also want to argue that South Africa can in a sense be seen as constituting a 'laboratory for modernity'. The local diffusion and expansion of the idea is integrally linked to the roles of a complex mix of actors, 'discourse coalitions,' that include donors, policy entrepreneurs, researchers, state actors and so on.

The Conference as networking agent in regional policy diffusion

Steiner-Khamsi argues that 'issue networks are institutionalised in the form of associations, journals, newsletters, list-serves and conferences. They export and expand the issue for their own particular survival.' (Steiner-Khamsi, 2004, p. 215) I have referred above of the research project as network agent. Another potent agent through which different actors interact and diffuse ideas is the conference. Mostly convened by donors, or with the assistance of donors, they act as lightning rods for the transmission, circulation, consolidation, legitimation, rejection and diffusion of particular ideas. Whether the ideas promoted in the official curriculum of the conference are accepted or not, the conference puts them into circulation

and ensures their absorption and wider dissemination. This does not occur in any simple way. There are different types of conferences, and they permit different types of discussions. They frame the discourse, and include and exclude what can be discussed. Whether the issues placed on the table are accepted or rejected, contested or not, they form part of an ongoing discussion that then reaches out more widely and touches other networks through the individuals who attend them. Always with an overt curriculum, the hidden curriculum of the conference becomes apparent in the issues that emerge outside of or in relation to the official agenda. Conference proceedings often crystallise these issues, bringing them into sharper focus.

This section focuses on two conferences, both of which produced proceedings, and both of which acted as a critical agent for the diffusion of ideas in the region (SAQA, 2003; Malawi National Examinations Board, 2005). The two conferences to be discussed are a conference hosted by the South African Qualifications Authority (SAQA) in 2003 in Pretoria on 'Qualifications and Standards: Harmonisation and Articulation Initiatives' and a conference hosted by a regional chapter of the International Assessment in Education Association association. The IIEA has an African chapter, the African Assessment in Education Association (AAEA). Three years ago, the Zambian Exams Board decided to host an assessment conference consisting of SADC bodies associated with AAEA. The 3rd Sub-Regional Conference on Assessment will be considered here. The Malawi National Examinations Board hosted it in Malawi on 'Educational Assessment in a Democracy: Challenges, Opportunities and Prospects.' In each it is important to consider the 'discourse coalition' (Schriewer, 2000b) of the conference, as well as the process of formation of the discourse itself (Schriewer, 2003) Through this, 'discourse coalitions' create particular discourses that are placed in circulation. This process of discourse formation gains power through use by significant actors and makes available discursive strategies for the formation of new policy and social action.

The 'discourse coalition' brought together by SAQA in 2003, in the wake of the Review of the NQF which had proposed a significant diminution of its authority, included the donor, DANIDA, a team from NQF structures and NGOs in South Africa, and consultants, trainers and others working in the field of qualifications and industry training from Australia, New Zealand, the United Kingdom, Mexico and representatives from 9 SADC countries. The aim of the Conference was 'to give form and content to the idea of a regional qualifications framework' through 'sharing experiences' and exploring 'ideas on the articulation and harmonisation of qualifications and standards globally, regionally and sub-regionally.'(vii) The Preface by the Danish ambassador acknowledged that 'setting up a regional qualifications framework will be challenging' and that the 'European experience in articulating qualifications has shown that it is a complex endeavour'. Its support of the Conference was essentially to support regional cooperation, 'dialogue, interchange and renewal.' The driving force in the Conference was the locally beleaguered South African Qualifications Authority. The motivation for development of a Regional Qualifications Framework embraced both the need to ensure greater mobility across borders as well as within contemporary systems of education.

A number of papers, including from the Mexican consultant and various industry-training consultants spoke for outcomes-based education and standards-generation. But it is clear

from the Editorial Review that 'behind the seminar and the process there was some controversy about the feasibility of a regional qualifications framework, and even some doubt about the nature and processes of national qualifications frameworks.' (p. 2) The Editorial Review thought that 'these doubts are fuelled by reports of teething problems and difficulties in various ambitious projects to run qualifications frameworks in some of the countries represented' and the view that 'qualifications frameworks have grown and solidified into complex regulatory systems with problems of workability and the effect of stifling what they aim to enable.' The Namibian paper expressed the classic rationale for national qualifications frameworks. It related the emergence of the Namibian qualifications framework to the discourse of global competitiveness and need for national economies to meet this challenge through human resource development. The NQF was seen as enabling labour market mobility and participation, lifelong learning and improved education and training through modernisation of its qualifications framework. Both the Namibian and South African presentations emphasised that learning from the experience meant ensuring that an NQF should be 'affordable, simple and meaningful.'(80) Like the Namibian presentation, the South African referred to the challenges at a technical level and at the level of 'stakeholder support,'(84) signalling that these had been in short supply. Both the Australian and the New Zealand presentations warned that, in the words of the Australian, 'the inevitable complexities of the system act as a barrier to other stakeholders.' (95)

The New Zealand presentation reported controversy around the introduction of unit standards across all sectors, that the process was costly, and that regional approaches in the South Pacific region were still fragmentary. Indeed, 'the philosophy implicitly (defining) the purposes of education in terms of the acquisition of units of competence …did not command respect and support across the full spectrum of educational and political opinion in New Zealand would be an understatement' (p. 116). Debate in New Zealand around unit-standards and competency-based education was 'acrimonious' 'and exacerbated the fault lines in New Zealand education' with 'much of the bitterness remain(ing) to this day.' (p. 116) This presentation made an impassioned plea that 'governments and quality assurance agencies should abandon the NQF model:

> Its conceptual and practical limitations are profound, and an NQF is thus likely to mislead, rather than inform labour markets, governments, students and other interested parties. This is not an argument for abandoning the educational and social goals that have often been identified with an NQF. Quite the contrary, the goals are of fundamental importance (access, equity, quality, skills and knowledge acquisition, human development) but an NQF is not the best means of helping to achieve those goals. Indeed, it may inhibit progress and be a positive hindrance. (p. 117)

Another South African paper, reflecting on the South African experience, noted the resistance in South Africa, too, and signalled the need to retreat on some fronts, but to build on the SADC front and to target in particular private and enterprise-related Technical and Vocational Education and training, 'the poorest corner of official structured learning in all SADC countries' (p. 25). The rationale presented for further work on qualifications frameworks was that anything other than an NQF was still part of a colonial inheritance.

Despite all the caveats and the retreats evident in all those countries that have taken on board National Qualifications Frameworks, the presentation on a SADC Qualifications

Framework for the region nonetheless was committed to the idea through the SADC protocol and on the basis of the idea of achieving a common system in the region. The view was that a qualifications framework could, as all the NQFs had thought, promote mobility and transfer across borders, whether national, institutional or personal. The idea is a great one, but the obstacles appear to be immense.

The conference provides an interesting example of how a discourse coalition is brought together and discourse formed around an idea to be exported: the narratives are those of ongoing work in the cause of a dream, of battles won and lost, of identification of the strategic points for continued work and intervention, and recommitment to an ideal. In this way, global, local and regional are brought together, visions and policies are promoted and ideas are circulated and made common currency, regardless of their usefulness or efficacy on the ground.

The second conference is completely separate from the first, and involves a different network. Different countries in the region host the annual sub-regional Conference on Assessment. It provides an opportunity for regional Ministry officials, university-based researchers, examinations commissions and others working in the field of assessment to talk about their work and exchange ideas. In 2003, for example, Zambia was the host and the conference theme was 'Learner Assessment for Improved Educational Quality.' Interestingly, the term 'learner' is a very specific South African-introduced reference to what elsewhere is referred to as 'pupils' or 'students'. In 2004, Zimbabwe hosted it with the theme: Benchmarking Ethno—based Learning and School-based Assessment in a Multi-ethnic and Multicultural environment.'

In 2005 Malawi hosted it with the theme 'Education Assessment in a Democracy: Challenges, Opportunities and Prospects.' Malawi's Second Development Plan for 1985-1995 stipulated a number of reforms. These include a shift towards 'methodology-centred and competency-based instruction.' One of the key proponents of South Africa's early form of outcomes-based education, whose work was roundly criticised for its incomprehensibility, was a consultant in Malawi on this shift in approach. Publishers followed soon after, looking for markets. (Interview Fathima Dada, Johannesburg, 24/5/2005) An entirely different network from the SAQA network, its papers also nonetheless reveal a preoccupation with and advocacy of modes of assessment associated with outcomes-based and learner-centred education. Increasingly, exams – deeply entrenched in systems in the region – are criticised and formative and continuous assessment are promoted. Outcomes-based education seems to have entered the discourse relatively uncritically with little real reflection on actual experience with it. To what extent it has actually been absorbed into and changed practice will require more classroom-based investigation than currently exists in the region. Indeed, while there is growing convergence at the level of the discourse, demonstrating interaction between people, there is little evidence of growing convergence at the level of practice.

Conferences are one of the main vehicles for the spread of new ideas and policies in the region. They provide a space for the discourse coalitions to be brought together to elaborate a common discourse and understanding of the role of qualifications frameworks and related institutions. But how and what may change in practice, and whether there is growing

convergence at the level of practice, requires an examination of actual country-level policies and practices. Such examinations in the vocational education and training arena reveal that there is little convergence at the national level, despite the fact that many countries have adopted qualifications frameworks and competency-based models of training. As one study concludes, 'systems (in the region) have evolved in a piecemeal and unsystematic way … that reflects historical accretions of institutions.' (McGrath, 2005, p. 8). In many instances, Atchoarena concludes, policy intentions have simply not been matched by implementation (Athcoarena, 2002). Given limited resources, and the evidence that many of these changes require resource-rich contexts, it is likely that practice in schools and classrooms has also not changed, and that the pattern is similar to that in vocational education and training, but much more work is needed to establish this and especially how school-based practices mediate and express the new discourses.

South Africa as a laboratory for modernity: the NQF

The idea that every universal idea has a local root, that policies are first experimented with in developing country contexts before being taken up in developed country contexts is an appealing one. The danger of the idea is the lurking conspiracy theory behind it and the way in which local agency is ignored. Thus, for example, we could perhaps argue that donors, consultants and the like imposed the National Qualifications Framework in South Africa and now in the region. National Qualifications Frameworks were embryonic in the UK and Scotland when taken up by South Africa. Consultants embroiled in the issues and debates in their own countries were certainly brought to South Africa to consult on how to bring such a framework into being. And donors also facilitated conferences, workshops and meetings between South African trade unionists and the developing competency-based movement in industry training in the developed world in the 1990s. However, the ideas and proposals brought in and facilitated by donors and consultants were eagerly sought after, worked upon and transformed by local practitioners.

As the South African National Qualifications Framework was brought into being, the international community watched to see what would unfold. South Africa could provide hope and new beginnings. And the NQF, not yet fully elaborated, or elaborated in the developed world, could provide a new beginning for the developed world. In 2004, the International Labour Organisation, which along with the World Bank, OECD and EU had facilitated and promoted South Africa's experiment, commissioned a British-based consultant who had been involved in the initiative from the beginning to review the experience of National Qualifications Frameworks (Young, 2004). The author notes towards the end that 'most of the experiences on which this report draws with the exception of South Africa come from industrialised countries such as England, Scotland and New Zealand – all countries with relatively high levels of resources.'(p. 60) And yet, the report draws on the differences of experience and lessons of implementation to reflect on why the South African experiment has been a failure. The main concern in the paper seems to be with South Africa. Issues central to the debates in South Africa about the NQF are lifted but dealt with by reference to the experience of other countries, except in some instances. It provides a damning critique of the experience of the NQF in South Africa.

The author locates the origins of the NQF in the competency-based frameworks, need for certification of youth employment schemes and shift from provider to employer-led training

that emerged in the neo-liberal environment of late 1980s Britain and New Zealand. It resurfaced in the mid-1990s linked to the idea of lifelong learning. The origins in Scotland were in universities. The author shows how most of the goals have not been realised in practice and how some of the assumptions have been fundamentally flawed. National Qualifications Frameworks promote credit accumulation and transfer through unit-based qualifications that are supposed to promote transportability and transparency and not require any specific prior learning programme. But in practice for example, he argues that credit-based qualifications have not provided for quicker progression up the educational ladder and qualifications frameworks cannot on their own do much to increase the demand for skills or knowledge and overcome deeply rooted inequalities that make for skill and educational differences. A basic fallacy of the notion of the NQF, he suggests, is the fallacy that education (or qualifications) can compensate for society (social inequalities). The author argues that most countries do have clear progression routes – for the majority the issues are not clear career paths; it is how to improve their capabilities to move up the ladder. NQFs, he argues, are top-down initiatives, using a grid-approach outside the realities and sites of formal institutions, curricula and pedagogies that creates problems because of their remoteness from practice. NQFs are designed independent of provision and are based on a single set of levels, standards and outcomes. An outcomes-based framework, he says, is essentially an assessment framework, not a framework for provision – and relatively few learners have the prior skills and knowledge to make use of such an assessment framework – most need provision as well.

The author distinguishes between different types of frameworks – enabling, strong and weak, partial and comprehensive, unit based and qualification-based frameworks and institution-led and outcomes-based frameworks. The starting assumption of most NQFs is that 'qualifications should be unit-based; in other words, the learning outcomes assumed to be necessary for a particular qualification are divided up into their basic elements or units.' (p. 24) The idea is that this approach maximises flexibility and choice for learners and employers. But in practice, here, too, he argues, it has created more problems than it has solved. The absence of a linkage between the qualification and provision is, he argues, a particular problem for developing countries: 'the danger is that qualifications will proliferate where there is no provision leading to them. An expensive activity without obvious wider benefits' (26).

What does he identify as the main problems with implementing NQFs? These cannot be resources or resistance, he says, as these characterise most reforms – they are principally political, administrative and technical or professional. Political difficulties are from the establishment of entities outside government with great power; administrative difficulties arise from 'the proliferation of new agencies and committees concerned with quality assurance, standard setting and assessment that NQFs invariably generate… They frequently have difficulty in recruiting members and staff with appropriate expertise (32). Technical or professional problems relate to the 'relatively straightforward activity of setting and marking examination, and for the need for new language of standards, units and levels to define criteria that have to apply to very different qualifications.' (33)

'Success stories', he says, have been incremental, based on trust, shared practice and policy breadth. The lesson from Scotland is that NQFs 'on their own have a relatively modest role

in reforming qualifications and improving vocational education.' (37) The lesson from New Zealand is that the establishment of powerful national organisations such as national qualifications authorities can lead to head-on confrontations with governments. The lesson from Ireland is to try not to be too prescriptive, but to emphasise at all times incrementalism, consensus and compromise.

The report is an interesting one for its metaphorical use of comparative perspective: it indirectly reports on and analyses the failure of the NQF, competency and outcomes-based models of education in the South African context by drawing on the other contexts in which NQFs have been implemented. It learns from the North to understand South Africa. And it helps the North and international agencies understand South Africa and its experiment – an experiment not to be supported in its current form. This is a report that is likely to influence ILO approaches to and support of NQFs in the region as much as in its own backyard.

Conclusion

The roots of new policy and curriculum approaches in southern and eastern Africa do not only lie in a borrowing process from South Africa, although particular NQF and outcomes-based aligned discourse coalitions in South Africans have tried hard to export their own brand of the National Qualifications Framework and outcomes-based education. Although South Africans borrowed and adapted the idea of a National Qualifications Framework themselves, South Africa was also a key test-site for the concept as a whole. Policy entrepreneurs have lent their expertise to South Africa and South Africans have lent theirs to the region. After a decade of experimentation, even those who initially consulted on it are now in serious doubt as to the South African model. Many who originally supported its development in South Africa are now pulling back. The paper has explored the conference as a vehicle through which discourse coalitions and policy entrepreneurs linked to the National Qualifications Framework and outcomes-based education create and circulate discourses, borrow and lend, selectively appropriate, modify and reject policies. Consultants are one form of policy entrepreneur, but these take many forms and include publishers, whose role has not been explored here. Regional conferences, in particular, as those discussed here exemplify, are a critical vehicle for diffusion of ideas and approaches on the embryonic Regional Qualifications Framework, as well as the approaches taken to outcomes-based education in the region. Whether these have had any bite in practice still remains to be seen, given the documented need for but absence of resources and capacity to enact these policies.

Acknowledgements

Thanks to Matseleng Allais for some of the sources.

References

- Allais, S. (2003) 'The National Qualifications Framework in South Africa: A Democratic Project trapped in a neo-liberal paradigm'. In *Journal of Education and Work*, 16, 305-324.
- Atchoarena, D. (2000) (Ed) The tranition of youth from school to work: Issues and policies. Paris: IIEP, UNESCO
- Atchoarena, D. and Delluc, A. (2002) (Eds) Revisiting Technical and Vocational Education in sub-Saharan Africa: An Update on Trends, Innovations and Challenges. Paris: IIEP, UNESCO.

- Bird, A. (2000) Youth and work in South Africa: Issues, Experiences and Ideas from a Young Democracy, in: D. Atchoarena (Ed) *The Transition of Youth from School to Work: Issues and Policies*. Paris: IIEP, UNESCO.
- Chisholm, L. Makwati, G., Marope, P. & Dumba-Safuli, S. (1998) *SADC Initiative in Education Policy Development, Planning and Management*. Harare: Canon Press.
- Chisholm, L. (2005) 'The Politics of Curriculum review and Revision in South Africa in regional context'. In *Compare*, 35 (1), 79-100.
- Crossley, M. & Watson, K. (2003) Comparative and international research in education: globalisation, context and difference. London & New York: RoutledgeFalmer.
- Dale, R. & Robertson, S. L. (2002) The varying effects of regional organisations as subjects of globalization of education. In *Comparative Education Review*, 46(1), 10–36.
- Daniel, J. Naidoo, V. and Naidu, S. (2003) The South Africans have Arrived: Post-apartheid corporate expansion into Africa, in: J. Daniel, A. Habib and R Southall (Eds) *State of the Nation 2003-2004*. Cape Town: HSRC Press.
- Daniel, J. Lutchman, J. and Naidu S. (2005) South Africa and Nigeria: Two Unequal Centres in a Periphery, in: J. Daniel, R. Southall and J. Lutchman (Eds) *State of the Nation South Africa 2004-2005*. Cape Town: HSRC Press.
- Department of Education (DoE) (2000, May 31) A South African curriculum for the twenty first century: report of the Review Committee on Curriculum 2005. Presented to the Minister of Education, Professor Kader Asmal, Pretoria.
- Department of Education and Department of Labour (2002) Report of the study team on the implementation of the National Qualifications Framework. Pretoria.
- Granovetter, M. (1983) 'The Strength of Weak Ties: A Network theory Revisited'. In *Sociological Theory*, 1, 201-233.
- Leyendecker, R. (2005) 'Curriculum, Examinations and Assessment in sub-Saharan Secondary Education (CASASE)': Early Draft presented at conference in Windhoek on 11 July 2005.
- McGrath, S. (2005) The multiple contexts of vocational education and training in southern Africa, in: Akoojee, S. Gewer, A. and McGrath, S. (Eds) *Vocational Education and Training in Southern Africa: A Comparative Study*. Pretoria: HSRC Research Monograph.
- Meyer, J. and Ramirez, F. (2003) The World Institutionalization of Education, in: J. Schriewer (Ed) *Discourse Formation in Comparative Education*., Frankfurt am Main: Peter Lang
- Ochs, K. and Phillips, D. (2004) Processes of Educational Borrowing in Historical Context, in: D. Phillips and K. Ochs (Eds) *Educational Policy Borrowing: Historical perspectives*. Cambridge: Cambridge University Press Symposium Books
- Pampallis, J. (2004) The Education Business: Private Contractors in Education, in: L. Chisholm (Ed) *Changing Class: Education and Social Change in Post-Apartheid South Africa*. Cape Town: HSRC Press.
- Sachikony, L. (2005) South Africa's quiet diplomacy: The Case of Zimbabwe, in: J. Daniel, R. Southall and J. Lutchman (Eds) *State of the Nation South Africa 2004-2005*. Cape Town: HSRC Press.
- Joel Samoff (1999) Institutionalizing International Influence, in: R. Arnove and C. Torres (Eds) *Comparative Education: The Dialectic of the Global and the Local* .Boulder: Rowman & Littlefield, pp. 51-89.
- J. Schriewer (2000a) World system and interrelation networks: The internationalization of education and the role of comparative inquiry, in: T. S. Popkewtiz (Ed) *Educational knowledge: changing relationships between the state, civil society and the educational community*. Albany: State University of New York Press, pp. 305-343.

- J. Schriewer (2000b) Educational Studies in Europe, in: E. S. Swing, J. Schriewer and F. Orivel (Eds) *Problems and Prospects in European Education.* Westport Connecticutt: Praeger.
- J. Schriewer (Ed) (2003) *Discourse Formation in Comparative Education*. Frankfurt am Main: Peter Lang.
- Spreen, C. (2004) Appropriating Borrowed Policies: Outcomes-Based Education in South Africa in G. Steiner-Khamsi (Ed) *The Global Politics of Educational Borrowing and Lending*. New York: Teachers College, Columbia.
- Steiner-Khamsi, G. (2003) 'Transferring Education, Displacing Reforms' in J. Schriewer (Ed) *Discourse Formation in Comparative Education.* Frankfurt am Main: Peter Lang
- Steiner-Khamsi, G. (2004) *The Global Politics of Educational Borrowing and Lending.* New York: Teachers College, Columbia
- Stoler, L. (2002) *Carnal Knowledge and Imperial Power: Race and the Intimate in Colonial Rule* (Berkeley, Los Angeles, London: University of California Press.
- Tyack, D. and Cuban, L. (1995) *Tinkering toward Utopia: A Century of Public School Reform* Cambridge: M.A., Harvard University Press.
- Vieira, S. Martin, W. and Wallerstein, I. (1992) *How fast the Wind? Southern Africa 1975-2000* Trenton, New Jersey: Africa World Press.
- Vinjevold, P. (2005) Presentation of Department of Education to IEB Assessment Conference, 24 August 2005.
- Watson, K. (2001) Comparative Educaitonal Research: The Need for Reconceptualision and Fresh Insights' in Watson, K. (ed) *Doing Comparative Education Research: Issues and Problems* Cambridge: Cambridge University Press Symposium Books.
- Young, M. (2004) National Qualifications Frameworks: Their Feasibility and Effective Implementation in Developing Countries. A Report prepared for the International Labour Organisation.

Standards based educational reform and its implications on school-based assessment

Lessons from schools

Francis Chirume, Zimbabwe School Examinations Council

Abstract

The question of assessment standards has been at the centre of debate at many regional and international conferences on education. The debates result from new education initiatives and call for increased accountability. This has raised the demand on schools to develop more effective, integrated methods for assessing student performance.

Expectations about what all students should learn—and, by implication, what they should be tested on—have changed in response to social, economic and technological changes and as a result of the standards based reform movement. According to the principles of standards based education, there are three general types of educational standards namely; content standards, performance standards and opportunity to learn standards; (Husen and Tuijnman, 1994). This paper presents the theoretical and practical aspects of standards based education and how these impact on school based assessment.

Zimbabwe secondary school teachers were interviewed on the effect of standards based education on school-based assessment. Respondents revealed that both content and performance standards are very critical in the implementation of school based assessment. The study established that teachers who are skilled in the area of assessment appreciate the idea of standards based education, and that it is essential and a necessary condition for the success of school based assessment.

It is recommended that teachers be in-serviced in the area of assessment and educational standards so that they can understand and apply the principles of standards in school based assessment.

Introduction

The issue of assessment standards has gained momentum at many regional and international conferences on education. Educational institutions are therefore held accountable for falling education standards. In Zimbabwe, the question of education standards, particularly assessment standards took centre stage in the Ministry of Education Sport and Culture from the early 2000. Policymakers reasoned that if schools and students were held accountable for student achievement, with real consequences for those that do not measure up to expectations, teachers and students would be motivated to improve performance. This has forced the government to introduce performance management and results based management in the civil service in general of which schools are part.

Educationists and policy makers have learnt from this decade of test-driven reform that testing can have powerful effects on teaching in classrooms. Ironically, those effects on classroom instruction, particularly when the test measures a narrow range of lower level skills, might narrow the curriculum and limit learning opportunities available to students. However, when standards of learning are high, and assessments are geared to such standards, student achievement may improve.

Current assessment trends (school based and problem based assessment) have necessitated examinations boards and Ministries of Education to identify higher standards for student learning and set content and performance standards that cannot be measured by low level tests. Indeed, with the advent of standards based reform, researchers, policymakers and education practitioners agree that methods for assessing student achievement must be revamped in order to better measure what students know and are able to do. This introduces the need for multi-dimensional assessment techniques that are all embedded in school-based assessment. The paper therefore focuses on the implications of standards based education on school-based assessment.

The paper presents an overview of standards based educational reform and issues involved in developing assessment to measure content and student performance standards. Teachers' responses on issues pertaining to standards based education and how they affect school based assessments are presented and discussed in the paper. Experiences are drawn from sampled secondary schools.

Research objectives

The objectives of the study were to:

1. Assess the applicability of the standards based education model in school based assessment
2. Assess the experiences of secondary school teachers in school based assessment
3. Establish the views and perceptions of teachers on the importance of
    i. Content standards in school based assessment
    ii. Opportunity to learn standards in school based assessment
    iii. Performance standards in school based assessment

4. Recommend the way forward for standards based education model in school based assessment.

Theoretical framework

The central purpose for schooling is at the core of the education system. The purpose is reflected by the question, "Who is expected to learn?". Traditionally, schools were designed to serve the best and the brightest, sorting out the others for manual labour jobs. Traditionally, and during the colonial era, white children (of high socio economic status) were expected to do best. The traditional educational system was designed to further that end. Standards based education introduces a different purpose, one that is focussed specifically on sustained student achievement for all students, regardless of gender, race/ethnicity, intellectual ability, or socio economic status.

However, designing and implementing a sustainable standard based system that consistently yields high student achievement involves more than setting and measuring academic goals. The process of changing to a 'standards' frame of reference goes much deeper. Changing to a standards-based system provides an opportunity to re-examine the organisational elements (Cordell and Waters, 1993) of a school system: fundamental purpose, principles, policies, processes, practices, programmes and procedures. Standards present an opportunity to examine or clarify these organisational elements as they are viewed by all the groups within the school community—teachers, school administrators, and parents.

In contrast to traditional beliefs, research has it that, although students do learn at different rates and in different ways, virtually all students can learn. Given appropriate time and instructions and clear expectations, many of the students previously written off can meet or exceed vigorous academic standards.

Standards-based education

The term 'standard' has been used synonymously to refer to curriculum standards, content standards and performance standards. Standards have come to mean many different things, to many different people. Kendall and Marzano (1995) distinguish curriculum standards, content standards, benchmarks and performance standards. Curriculum standards, they explain "are best characterised as descriptions of what should take place in the classroom, as such, they address instructional techniques, recommended activities, and various modes of presentation (p. 20). Content standards describe what students should know or be able to do. Performance standards refer to the quality of the performance deemed acceptable for each content standard.

According to Husen and Tuijnman (1994), the standards-based educational reform has three general types of educational standards that include: content standards, performance standards, and opportunity-to-learn (OTL) standards. Each type has an indispensable part to play in the ideal model of standards based reform as highlighted below.

Content standards

Content standards describe the range of desirable knowledge and skills within a subject area. They specify the general domains of knowledge that students should learn. Mueller (2004)

views content standards as statements that describe what students should know or be able to do within the content of a specific discipline or at the intersection of two or more disciplines. Examples of content standards would include:

- Students will classify objects along two dimensions
- Describe effects of physical activity on the body
- Present employment-related information in the target language.
- Solve three-dimensional problems involving the angle between a line and a plane.

In History for example, content standards specify the people, events, and ideas that should be included in the history curriculum at each grade level.

Performance standards

Linn, Baker and Dunbar (1991) define performance standards as specifications of 'how much' students should know and be able to do. Thus, while content standards shape what goes into a curriculum, performance standards set benchmarks, that is, specified levels of achievement that shape expectations for educational outcomes provide a basis for measuring learning outcomes and provide the criteria for imposing rewards and sanctions. Performance standards are therefore statements that describe skills students should develop to enhance the process of learning. They are generic skills that students should display for specified levels of achievement.

Performance standards for Mathematics for example, specify the mathematical operations and concepts that should be mastered at each grade level as well as the types of assessments that should be used to measure that mastery.

Opportunity-to-learn (OTL) standards

Opportunity to learn (OTL) standards specify the nature of educational inputs and resources that are needed to realise expectations for student and school performance, (Stites 2004). Stites (2004) further maintains that OTL standards are needed to respond to concerns over the potential inequity of raising expectations for all students without ensuring that all had an equal opportunity to meet higher expectations. For example OTL standards might specify the number of hours and quality of instruction that students should receive before they are tested on desired levels of skills and knowledge specified in content and performance standards.

The three types of standards (content, performance and OTL) are linked and interrelated. For meaningful and fair performance standards to be set, it is necessary to define the exact content areas to which these standards shall apply. Before performance can be fairly assessed, it is moreover necessary to determine whether all students have had adequate opportunities to learn the prescribed content (Hussen and Tuijnman 1994). In fact OTL standards should be predetermined and set first before content and performance standards; as these form the basis for comparability purposes.

It is clear that performance standards carry most of the load in the theoretical model. They specify 'how much' students should know and be able to do. They may also specify the tests used to measure whether students have learned enough. For these reasons, performance

standards supply the leverage needed to hold learners and educational programmes 'accountable' for learning. While content standards may be the result of broadly inclusive efforts to achieve consensus on 'what' students should know, in the end, they simply describe what 'ought to be' not what 'must be'. Content standards alone do not drive a system of accountability for educational outcomes. Performance standards, with accompanying indicators to specify where to look to see how much is there and benchmarks to determine the level of performance that is enough, are considered clear performance standards.

In the ideal model of standards-based educational reform, content, performance and OTL standards each have clear and distinct roles to play. For example, for performance standards to be fair, students and others who will be held accountable for outcomes must have the opportunity to meet those standards, therefore there should be standards for the quality of schools and schooling.

School-based assesssment

Several authors have given varied definitions of the concept school-based assessment. Ogunniyi (1984) views school-based assessment as a formative evaluation procedure concerned with the finding out, in a systematic manner, the overall gains that a student has made in terms of knowledge, attitudes and skills after a given set of learning experiences. The main idea is that school-based assessment should evaluate all the domains of learning, that is:

- Cognitive domain (acquisition of knowledge)
- Affective domain (beliefs, attitudes and perceptions)
- Psychomotor domain (acquisition of practical skills).

In other words SBA should evaluate the total human being. School-based assessment is basically concerned with the progress of pupils by regularly measuring their performance against curricular goals and making the results available to classroom teachers. The result of that assessment can be used as part of the scoring of public examinations. Alternatively, they could become the sole basis for certification or selection. What then are the key characteristics/features of school-based assessment?

Characteristics of school-based assessment

School-based assessment is conceptualised as an on-going assessment exercise in schools. From its definition SBA has the following features as defined by Obioma (1986):

- **Comprehensiveness:** School-based assessment focuses on all the three domains of learning, making use of a variety of measurement instruments.
- **Cumulative nature:** In school-based assessment, the previous assessments of the learner are taken into account in determining future achievements in a given task or course of study.
- **Systematic procedures:** School-based assessment depends on a well-developed operational plan where the intervals of measurement, assessment and evaluation of the learner's achievements are clearly stated.

- **Guidance oriented:** School-based assessment is diagnostic and has a feedback mechanism, which permits the application of some remediation and intervention.

These features constitute an effective school-based assessment programme.

Methodology

The study population comprised secondary schools in the nine educational regions of Zimbabwe. Out of the nine regions two were purposefully selected, that is Harare, and Mashonaland West Region. The two regions were selected on the basis that one region (Harare) had its teachers trained in assessment principles while the other one (Mashonaland West) had not been trained in assessment. The Zimbabwe School Examinations Council trained Harare secondary school teachers in the following areas of assessment:

- The syllabus and its components (syllabus interpretation)
- Forms of assessment
- Test Construction
- Marking
- Analysis of classroom tests.

Training was done so that teachers could incorporate skills gained in their formative evaluation of pupils, particularly school-based assessment.

Structured interviews were administered to a sample of 60 teachers constituting 30 per selected region, while focus group discussions were held with teachers where interviews had been conducted. Interviews complimented with focus group discussions, gave the researcher an insight into how teachers perceive standards based education, namely content, performance and opportunity to learn standards on effective school based assessment. Data was both qualitatively and quantitatively analysed. Quantitative analysis of data resulted in percentage responses being computed to give more meaning to teachers' responses.

Discussion of findings

Study findings are presented and discussed according to themes to which they fall in. The findings are discussed according to themes such as, content standards and school based assessment, opportunity to learn standards in school based assessment, and performance standards in school based assessment. Views of teachers who had been trained or not trained in assessment were discussed under these sub-themes.

Content standards and school based assessment

Content standards describe what students should know or be able to do, and these are captured in National Syllabuses. The Curriculum Development Unit (CDU) in consultation with key stakeholders develops all syllabuses in Zimbabwe. The syllabuses combine both curriculum and content standards.

Teachers are supposed to analyse the syllabus and conceptualise all the curriculum and content standards during teaching. About 80% of the teachers who had been trained in syllabus interpretation indicated that they used the syllabus to plan and align content

standards to their schemes of work. They set school content standards on the basis of syllabus content standards. The school syllabuses analysed captured all the curriculum and content standards in the official syllabuses. School based tests are constructed from content covered in the school syllabuses.

Harare teachers also indicated that members of the school reach consensus on content standards, and determine what is learned at each organisational level. In addition, a variety of materials are used which provide experiences aligned with content standards. Teachers then teach to content standards.

On the other hand, 75% of teachers in Mashonaland West Region indicated that their scheming and planning was based mainly on prescribed textbooks. Teachers in this region had not been trained in syllabus interpretation and how to analyse important sections of the syllabus. The majority of teachers in this region said that they only used the syllabuses simply to check topics that should be included in the schemes of work. The breadth and depth of content to be covered were taken from recommended textbooks.

It must be pointed out that textbooks though they are prescribed or recommended do not necessarily cover all the content as prescribed in the official syllabuses. In addition textbooks remain static while knowledge changes with time. It can be concluded that teachers who use textbooks to prepare schemes of work for teaching or to prepare school syllabuses may fail to align school content standards to national content standards. If school content standards do not match national content standards then this would have a negative impact on school-based assessment since classroom tests are constructed on the basis of content taught. This heavily affects the quality of school-based tests.

Opportunity to learn standards and school based assessment

Opportunity to learn standards (OTL) are also known as delivery standards. They include educational resources and materials necessary for effective teaching and learning. The rationale for OTL standards are hinged on the idea that, for performance standards to be fair, students and others who will be held accountable for outcomes must have the opportunity to meet those standards, and therefore there should be standards for the quality of schools and schooling.

An analysis of national syllabuses for secondary school subjects revealed that all the syllabuses do not indicate opportunity to learn standards. Opportunity to learn standards should provide all the candidates who are at the same level at least equal opportunities for them to learn. Results of focus group discussions with both groups of teachers indicated that opportunity to learn standards are critical if candidates are to be compared and measured along the same continuum. Respondents revealed the following opportunity to learn standards as critical in school-based assessment:

- Equal instruction time for all students for a given subject
- An equitable distribution of educational resources; such as textbooks, equipment and stationery
- Standard learning environments
- Standard learning facilities such as furniture, buildings and grounds

- Well-qualified teachers
- Varied and student centred teaching/learning methodologies.

All these aspects of opportunity to learn standards are not captured and specified in the syllabuses with the exception of teaching methodology, but however the Ministry of Education Sport and Culture attempts to ensure equity in the distribution of such resources. Standard instruction time for each subject is specified by the Ministry as shown in the table below.

Instructional time standards for selected subjects

Level	Subject	No. Of Periods per Week
ZJC	Mathematics	6
ZJC	Shona	4
ZJC	History	4
"O" level	Mathematics	6
"O" level	Science	6
"O" level	Shona	4
"A" level	Physics	12
"A" level	Maths	12
"A" level	Geography	8

All the teachers interviewed reported that they implement recommended instructional time standards. Therefore recommended teaching time standards were followed in schools in the two sampled regions. However, there were variations in the distribution of educational resources such as textbooks, equipment and even qualified teachers. Disparities in learning facilities such as furniture, buildings and even sporting grounds were noted between urban schools and rural schools. Urban secondary schools enjoyed better facilities than rural secondary schools. Teachers were of the view that opportunity to learn standards must be uniform across all schools if school based assessment is to be effectively implemented and contribute to national examinations. However, close to 65% of the trained teachers echoed that disparities in opportunity to learn standards would continue simply because of the differences in administration and priorities that are attached to by each School Development Committee/Association. Differences in these standards would mean that school based assessment standards would differ from one institution to another with schools that enjoy better resources and facilities set to perform better than those with inadequate resources. These would basically affect the quality of school-based assessments for poor secondary schools. From the teachers' responses there is a link between opportunity to learn standards and effective school based assessment. Poorly set opportunity to learn standards would result in ad-hoc school based assessments.

It has been mentioned that delivery standards (teaching methodology) are part of opportunity to learn standards. These are very critical in school-based assessment. All the teachers interviewed responded that instruction should be varied. They indicated that teachers prepared a variety of approaches that reach all students. Teachers who had been trained in school based assessment were of the opinion that multi-dimensional teaching approaches should be applied if standards are to be realised in school based assessment. The multi-dimensional approaches should encompass ethno-based methodologies, which are

very critical in school-based assessment. Ethno-based methodologies are naturalistic approaches that make maximum use of the environment in teaching and learning. It was noted that 82% of the teachers who had been trained in syllabus interpretation, teaching methodologies and test construction appreciated the application of delivery standards to effective teaching and learning. This observation concurred with perceptions of 20% of respondents in Mashonaland west region who had not been trained in the above aspects. The implication is that the remaining 80% of the teachers who had not been trained together with 18% of those who had been trained need staff development on ethno-based teaching/learning so as to improve instruction in schools. In short, capacity building programmes on modern teaching approaches should be planned for classroom teachers.

Performance standards and school based assessment

Classroom assessment is most effective if what gets taught gets tested; if classroom assessment is aligned with national level accountability assessment and content standards; and if all these are congruent with purpose, principles, policy and practice. Aligned classroom assessment enables the teachers to make instructional decisions for students on continual basis. Classroom assessments allow students to practice skills from simple to complex and to integrate those skills in meaningful ways. Students must know what skills they currently have and what they are expected to do in order to meet or exceed the standards.

Performance standards are specifications of how much students should know and be able to do. Performance standards include assessment standards. If assessment standards are to be met in schools, teachers should have knowledge of assessment principles and display skills and competencies in:

- Test construction, that is construction of various types of tests which include multiple choice, structured, essay, practical, projects and portfolios
- Marking
- Test analysis
- Grading and Awarding
- Reporting pupil performance.

Harare teachers were trained in assessment principles and procedures of test construction, marking and analysis of classroom tests while Mashonaland West teachers had not been trained. A comparative analysis of responses from these groups of teachers on performance standards as they are applied to school based assessment is highlighted.

Document analysis of syllabuses revealed that the syllabuses lacked detail of performance standards. Most of the syllabuses include assessment objectives and scheme of assessment with a few indicating the specification grid. Performance standards however do not only include these three aspects that guide assessment but also incorporate grade descriptors as indicators of performance. Grade descriptors indicate the generic skills that candidates should display at each grade level. There is need therefore to include grade descriptors (with requisite generic skills) in all national syllabuses so that schools would make use of them and align their school-based assessments to these generic skills.

A sample of teachers who had been trained in assessment were able to set assessment standards for their institutions, a move which impacts positively on school based assessment. The teachers indicated their ability to construct various forms of tests, which they administered in their institutions. On the other hand close to 25% of the teachers who had not been trained in assessment appreciated the use of multi-dimensional assessment techniques (portfolios, performance assessment and authentic assessment) as current trends in student assessment in school based assessment. This was so because multiple choice formats and tests that require the use of paper and pencil only have been criticised because they do not reflect practices in the 'real world', (Burger 1998). Therefore performance and assessment standards should be included in the national syllabuses so that teachers can align their school performance standards to these national benchmarks. There is need also to train teachers in all aspects of assessment in order to improve school based classroom tests.

Conclusions and recommendations

From the discussion, it was noted that in the standards based educational reform model, all forms of standards, that is content, performance, and opportunity to learn standards should be aligned. Teachers can apply the model in teaching so as to improve results. This will help in benchmarking teaching/learning and assessment standards to national standards. First, coherent content standards can provide a clear vision of what every student should know and be able to do. Performance standards and related assessment will provide the tools for individual learners and everyone else to monitor progress toward goals. The educational standards movement can be characterised as an assessment driven reform effort. As such, the basic idea here is that since teachers often teach to the test, one way to improve teaching and learning is to create a better test, leading to authentic school based assessments.

Results of the study indicated that teachers who had been equipped in the area of assessment, used the syllabus regularly in their teaching while those not trained used recommended textbooks. Therefore content standards were compromised in these schools. The use of textbooks fails to determine the breadth and depth of content to be covered. The study revealed disparities in the allocation of educational resources. This meant opportunity to learn standards varied from one institution to another, resulting in urban schools having better resources than rural schools. Opportunity to learn standards were not indicated in the syllabuses with only the exception of teaching methodologies. There is need to specify opportunity to learn standards for effective school based assessment.

In the area of performance standards, it was revealed that syllabuses lacked detail on performance and assessment standards. Although the syllabuses contained assessment objectives, scheme of assessment and the specification grid, they did not have grade descriptors, which indicate the generic skills candidates should display at each grade level. Teachers therefore could not align school-based assessments to national standards. In the same vein, teachers trained in test construction showed the ability to prepare various types of tests as opposed to those not trained, a move that has a positive impact on the quality of school based assessments.

On the basis of these findings, the following recommendations were made:

- Schools should implement the standards based educational model in their teaching /learning and assessment. This would enable schools to be accountable and be in a position to align their standards.
- The national syllabus should be the official teaching document where teachers should derive their content standards.
- National syllabuses should capture opportunity to learn standards as well as performance standards (especially grade descriptors) so that teachers can align their school based assessments.
- Secondary school teachers should be in-serviced in ethno-based teaching/learning, assessment (syllabus interpretation, test construction, marking and test analysis) and school-based assessment.
- Capacity building programmes on assessment and ethno-based teaching/learning should be extended to teacher education colleges and universities that train teachers.

References

Burger, D. (1998) Designing a Sustainable Standards-based Assessment Systems. Mid-continental Regional Educational Laboratory.

Cordell, F. D., and Waters, J.T. (1993). Improving Student Performance: New Strategies for Implementing Higher Standards. Greeley, Co: The Centre for Peak Performing Schools.

Husen, T., and Tuijnman, A. (1994) *Monitoring Standards in Education: Why and how it came about*. New York: Pergamon.

Kendall, J. S., and Marzano, R. J. (1995). The Systematic Identification and Artculation of Content Standards and Benchmarks. Aurora, Co: Mid-Continent Regional Educational Laboratory.

Linn, R.L., Baker, E.L. and Dunba, S,B. (1991). 'Complex performance based assessments: Expectations and validation criteria'. In *Educational Researcher*, 20(8) 15-21.

Obioma, G. (1986) Introducing Continuous Assessment in A. Ali (ed). *Measurement and Evaluation*. Ibadan: Heinemann.

Ogunniyi, M.B. (1984) *Educational Measurement and Evaluation*. Lagos: Longman

A recognition of outcomes-based assessment practices from a Foucauldian perspective

Cindy Ramhurry, University of Johannesburg

Abstract

Assessment in *Curriculum 2005* is premised on a progressive learner-centred, outcomes-based approach where observation, monitoring and surveillance are key processes advocated for success in learning (NDE, 1997). This paper argues that despite the seeming liberatory nature of outcomes-based assessment, optimism is this regard must be mediated by an understanding that 'invading privacy' may be one of the most disturbing aspects of the above-mentioned processes at work. The paper highlights some of the tensions evident in outcomes-based assessment practices in a well resourced senior South African school. An exemplary situation that captures this argument is drawn from my doctoral study and analysed through a Foucauldian lens wherein he forges a connection between increasing surveillance, on the one hand, and escalating control and regulation, on the other (Foucault; 1998). The paper is guided by the question: What forms of power and control are carried within the new forms of assessment? The literature review focuses on relevant concepts of Foucault's work to highlight the problematic nature of outcomes-based assessment for learners. Several implications emanating from the analysis for assessment in South Africa are then examined and discussed.

Introduction

The significance of this paper lies in its contribution toward a theory of power relations in assessment practices which, in addition to extending past macro-level theories, will account for patterns in the micro-level circulation of power and disciplining of bodies in South Africa. Moreover, this study contributes to such theory from a unique empirical base. To the extent that power relations are central in the micro-level enactment of pedagogy, and that enormous continuities over time are evident in the character of what occurs among teachers and students, a theory of power relations in pedagogy which identifies patterns and specific practices in such a way that enables new points of intervention to be explored, has the

potential to make a substantial contribution not only to educational theory, policy, and practice but also to social theory more generally.

The relevance and pervasiveness of postmodernism in most spheres of human existence is such that it should not be ignored in educational assessment practices (Hargreaves, 2002). The application of one of these postmodernist doctrines and its practice to educational assessment practices is central to this paper, which justifies its use as a point of entry into the discussion. The paper generates the problematic of power in assessment practices, employing a theoretical framework which is grounded in a conception of power based on the work of Michel Foucault (1977). In this paper I argue that outcomes-based assessment in some respects epitomizes progressive educational themes yet, when studied with an eye toward productive power reveals several contradictions and paradoxes. I support my argument with empirical evidence drawn from my doctoral study conducted in a grade 8 class-room in a South African secondary school. The Foucauldian tools of analysis are then described and applied to an exemplary assessment situation in order to demonstrate the technologies of disciplinary power within the new forms of assessment and what this technology does to teachers and learners when they become involved in it. The implications of these practices of power are then discussed.

Postmodernism, inter-subjectivity and assessment

Postmodernism and its ancillary wave of radical changes such as globalization and progress in information technology have called for a fundamental paradigm shift in the philosophy, structure, and contents of educational policies worldwide (Mockler, 2004). Assessment policy reform has been a common trend on both international and local levels resulting in significant implications for practice. The key principle underpinning the shift is the move from the 'objective' to 'inter-subjective' perspective of assessment. In traditional processes of assessment; the assessor (whether a teacher, researcher or stranger administering a group or individual test) is seen as a detached 'eye' and views the testing event from an objective distance, outside the universe of the person assessed. That distance assures the objectivity of the score. This perspective has been criticized on account of its preoccupation with measurement of rather than the learning itself. Postmodernists advocate an 'inter-subjective' perspective which is believed to provide greater empowerment in learning. Bruner argues that inter-subjectivity is

> the human ability to understand the minds of others, whether through language, gesture or other means. It is not just words that make this possible, but our capacity to grasp the role of the settings in which words, acts and gestures occur. We are the inter-subjective species par excellence. It is this that permits us to "negotiate" meanings when words go astray (Bruner 1996: 20)

Outcomes-based assessment policies introduced in South Africa have a post-modernist approach: they are characterised by collaborative and self-directed inquiry using democratic and dialogical assessment strategies such as self-, peer- and group- assessment strategies (DoE, 2000). These policies embody 'inter-subjectivity' in their quest to foster empowerment at the 'micro' level of bodies, through students becoming self-regulating and active participants in their own learning.

Almost all questions in educational assessment it would seem are more or less connected with the issue of power. In fact, "Power has become one of the central concepts of the social and human sciences *per se*" (Clegg 1989, xviii). In terms of existing empirical studies of pedagogy, while there is a vast body of literature on pedagogy, there is little empirical work that attends both to issues of pedagogy and to issues of power. Although Bourdieu (in his work with Passeron and others) and Bernstein provide the most sophisticated and detailed analyses of power and pedagogy, even their work on this topic lacks a substantial systematic empirical base (Gore, 2002). Bernstein's (1990) theory of power (related to Marx) is derived from mapping societal power relations and showing that pedagogy is implicated in the production of those relations through its rules (regulative, distributive, contextualising, etc.). Bourdieu and Passeron's (1977) view of pedagogy as symbolic violence is based on Weberian notions of authority and the emphasis is on how large discourses are imposed on, and taken up by, the body (Gore, 2002).

My own view, derived from Foucault, goes beyond Bernstein's (1990) declaration that pedagogy is more than a relay for power relations external to itself to *demonstrate* how the power relations inherent to pedagogy govern and regulate bodies and knowledge and to show that discourses are constructed out of pedagogy itself. From my perspective, power is not simply the imposition of one will on another. Its subtleties and nuances are taken into account in a way that acknowledges the complexity of classrooms in which there is much more going on than the imposition or reproduction of broader societal power relations. In terms of existing studies of assessment, while there is a vast body of literature that concentrates on macro-practices of assessment, very little energy is dedicated to the micro-level of analysis. The study from which this paper is drawn is designed to bridge that gap by examining micro-practices of power in classroom assessment practices.

Classroom assessment practices are therefore regarded as both central and peripheral to this paper. The unveilling of technologies of power in classroom assessment practices occurs, here, on two levels: locally, in a study of outcomes-based assessment in classrooms, and discursively, on the value of a Foucauldian analysis of power for educators. The two stories work concurrently. Foucault's analysis of productive power provides the theoretical grounding for a study of power operating in assessment practices. In turn, the study of power in assessment practices serves to illustrate gaps in the predominant ways power is addressed in outcomes-based assessment contexts. Rather than offer a full-scale genealogy of classroom assessment practices, I contribute a genealogical snapshot of one particular transition in teaching practice, namely the shift from so-called 'traditional' assessment to 'outcomes-based' assessment frameworks of organization.

Background to the study

The site for the broader study was Suburban High School (pseudonym); a grade 8-12 building located in the suburbs of a very large, metropolitan area with a human population numbering in the millions. The observations at Suburban High School were conducted within teachers' classrooms in the subject areas of Mathematics, English Main Language, and Arts and Culture. The key participants in this study were three teachers in the learning areas of Maths, Languages (English Home Language) and Arts and Culture and their respective learners. Although a small sample limited any possibility for generalizability, this case study research dug deeper and looked more broadly than would be conceivable with some

commonly used quantitative methods. Data reduced from interview transcripts and video-recordings resulted in thematic perception generalizations toward 'inter-subjective' assessment practices at Suburban High School. Although the broader study explores more fully the concept of 'inter-subjective' assessment practices, for the purposes of this paper only one of these elements, namely, collaborative learning practices is given priority. Therefore, this paper looks closely at the configurations of power inherent in peer-assessment practices and the kinds of subjects it produces. The question that guided this aspect of the research was: How are patterns of power currently determining the students' experiences of assessment, and constructing students' image of self, others, and the world?

Theoretical framework

As discussed earlier, the theoretical framework for the paper is grounded in a conception of power based on the work of Michel Foucault (1977). Key features of this conception of power are that power is productive and not solely repressive, that it circulates rather than being possessed, that it exists in action, functioning at the level of the body, and that it operates through 'technologies of self', that is, that individuals are active in their own subjection. Foucault's conception of power (power relations) requires a focus on the mechanisms of pedagogy rather than on individuals or groups who might traditionally have been seen as holding power. Hence, the primary research question was "What specific practices actualise relations of power in assessment?" While other scholars and researchers in education have engaged with some of Foucault's ideas (e.g., Ball, 1990; Cherryholmes, 1988; Jones and Williamson, 1979; Marshall, 1990; Meredyth and Tyler, 1993; Walkerdine, 1990), this paper aims to demonstrate, rather than assert, the applicability of Foucault's thought to the study of power.

Foucault's concept of disciplinary power explicitly shifts analyses of power from the 'macro' realm of structures and ideologies to the 'micro' level of bodies. He argued that unlike the sovereign power of earlier periods, disciplinary power functions at the level of the body:

> In thinking of the mechanisms of power, I am thinking rather of its capillary form of existence, the point where power reaches into the very grain of individuals, touches their bodies and inserts itself into their action and attitudes, their discourses, learning processes and everyday lives. (Foucault, 1980, p.39) (emphasis added)

Foucault (1980) elaborates the invisibility and pervasiveness of power in modern society: "The eighteenth century invented, so to speak, a synaptic regime of power, a regime of its exercise within the social body rather than from above it" (p.39). Using the exemplar of the Panopticon, with its normalising surveillance, Foucault described disciplinary power as circulating rather than being possessed, productive and not necessarily repressive, existing in action, functioning at the level of the body, often operating through 'technologies of self'.

Of significant relevance to the strategy of collaborative assessment practices is the metaphor of the panopticon. Perhaps the most-discussed section of Michel Foucault's Discipline and Punish is his description of Jeremy Bentham's panopticon, a system of surveillance originally designed with penal institutions in mind, but that has become a metaphor for the much broader and more subtle intrusion of observation and record-keeping techniques into more and more areas of social life (Foucault 1977: 200-209); see also Poster (1990) Foucault

(1980); Bourdieu and Passeron (1977) Bernstein (1990) Bernstein's (1990) Passeron (1977: 69-98). The basic idea of the panopticon is straightforward: a central tower or structure has windows on all sides, and it is in turn surrounded by a ring of cells occupied by the inmates, the open sides facing inward. Observers can look out in any direction, at any time, to see what any inmate might be doing. Furthermore, since the inmates cannot see into the central observation tower, every window or observation point does not, in fact, need to be staffed all the time; the possibility of being observed has a deterrent effect even when inmates are not in fact being observed. And, still further, as inmates become accustomed to this environment, and to the routine of assuming that they are being observed at any/every time, it becomes less important for the observation tower to be staffed at all; the structure of the environment is what exerts control, as people internalize changes to their habits and movements without remembering the original circumstances that necessitated them. The panoptic condition becomes part of the identity of an inmate ("a madman, a patient, a condemned man, a worker or a schoolboy" (Foucault, 1977: 200)).

Several deeper conclusions for this study follow from Foucault's discussion. The first is that such mechanisms of surveillance tend to become more pervasive: for example, few people even notice any longer how frequently they are monitored through partially hidden video cameras (from the bus, to the bank, to the store, to the parking lot, to the elevator). This is one of the central themes of Foucault's book: that as the mechanisms of surveillance and control become more subtle and 'humane', they become more extensive; they actually become more controlling in their effects, but with less complaint.

In addition, I drew on the work of Gore (2002), who has developed a descriptive theory of the way pedagogy functions in classrooms from Foucault's analysis of power and existing studies of pedagogy. The following working definitions were adapted from the work of Gore (2002) in order to cover a wide range of micro-techniques, or practices, of power in assessment.

- **Surveillance:** Supervising, closely observing, watching, threatening to watch, avoiding being watched
- **Normalisation**: Invoking, requiring, setting or conforming to a standard, defining the normal
- **Exclusion**: Tracing the limits that will define difference, boundary, zone, defining the pathological
- **Distribution**: Dividing into parts, arranging, ranking bodies in space
- **Classification**: Differentiating individuals and/or groups from one another
- **Individualisation**: Giving individual character to, specifying an individual
- **Totalisation**: Giving collective character to, specifying a collectivity/total, will to conform
- **Regulation**: Controlling by rule, subject to restrictions; adapt to requirements; act of invoking a rule, including sanction, reward, punishment
- **Self(/r/t/s)**:Techniques/practices directed at the self by researcher, teacher or student

Using these theoretical tools, I conducted multiple analyses of data, including the application of two sets of coding categories to the data—the first set designed to identify techniques of power derived from Foucault and a second set emerging from a grounded analysis of the data. In this paper, however, the analysis is limited to findings emerging from the first coding process.

Given clear patterns evident from the qualitative analysis of these data, an attempt was made to quantify the results in order to identify patterns in the proportion of each practice of power relative to the other practices within subjects, and to make comparisons between subjects. Furthermore, each coded segment has been analysed, in a more contextualised way, for the object of the particular practice of power, the specific way in which the technique of power was enacted, the direction of the exercise of power, and any reactions or consequences evident. Sites in which reflected a variety and multiplicity of codes were marked as 'critical incidents' for further comparison and analysis.

Background to 'critical' incident

The incident is drawn from a grade 8 English Language lesson in which, 'clauses' in sentences is assessed through a collaborative group experience. The incident captures power issues in the build-up to main assessment activity in which pupils are put into groups and assessed on their performance of a task. The teacher, Jenny introduced the topic by building on a section she started in the previous lesson. She used varied methods to disseminate information to pupils including 'chalk and talk', the use of the overhead projector and the distribution of worksheets to pupils. Pupils were seated in traditional rows (one behind the other) for part of the lesson and then moved into circular formations for the group-activity. Jenny positioned herself at different monitoring points during different stages of the activity. Prior to this critical incident Jenny had explained what a clause is and provided various examples thereof. The students were then asked to copy some of these explanations and examples from an overhead transparency. Jenny then embarked on a process of distributing students into groups for the peer-activity. Students were then engaged in group-work. During the activity, which became rowdy, Jenny repeatedly asked for better behaviour from the class. These requests were polite at first and then became firm commands. She then outlined a detailed procedure in which each group was given a specific "focus-group" for the peer-assessment. The final stage of the activity culminated in the performance of the required task in the front of the class and the peer-awarding of marks on a rubric.

On a structural level of analysis, almost midway through the lesson and also at a very strategic point (the critical incident at which the assessment was introduced), the entire ethos of the lesson dramatically changed. The classroom became a flurry of activity with an engagement in learner-centered activity involving much group-work and peer-group assessments. The teacher's role appeared to change radically and she walked around the class as opposed to her 'objective' stance in the earlier half. Her position of observation took on an 'inter-subjective' form (Bruner, 1996). The teacher handed out 'rubrics' to learners. It was significant to note that although there was much query on the 'rubric' in use, not much information was forthcoming from the teacher. She reminded pupils that their marks would be recorded. The analysis reflected a classic case of the lesson splitting neatly in two as the assessment came in. It could almost be perceived as though the assessment was perceived as an intrusion into her classroom. Questions of why a very strongly structured and planned

lesson suddenly loosened up into this strange amalgam of peer assessment was found to be weird in the extreme and provided an immediate avenue for further exploration. An attempt is account for this question further in the analysis.

Discussion of findings

Initial levels of analysis revealed that there were comparatively high levels of surveillance, regulation and distribution in terms of the exercises of power by the teacher in relation to the students. Classification was used by both teachers and students. When used by teachers it was used nearly as often for disciplinary purposes as it was for instructional purposes.

Surveillance

In the site, surveillance was mainly exercised by the teacher in relation to the students and evident primarily in relation to classroom management. Surveillance was found to be strongly linked with regulation. However, there was some surveillance of the teacher by the students; student surveillance of the teacher took the form of watching to ensure the teacher wasn't watching their own (unacceptable) behaviours. Surveillance took both visible and articulated forms. Surveillance also was used as a pedagogical technique by having students watch each other's performance (physical skills, role plays). Objects of surveillance included students' movement (in and out of the classroom, their placement in groups, their participation and objects they touch (especially objects they should not touch), their activities, class-work (progress, assignments, books) and behaviour. In the site, surveillance was mainly used to gain student attention but rather than using threats or disciplinary action, more subtle techniques such as making eye contact and circulating among students were used. There was also formally sanctioned surveillance or students' attendance and of their performance on the task (in a teacher's journal). Students were constantly under surveillance in relation to behaviour (noise, restlessness, inattentiveness).

Regulation

In the site, regulation primarily centred on student behaviour and was directed from teacher to students. It was also found that regulation was more concerned with assessment requirements and negotiated class rules. The recording of information in a personal journal may also be seen as a formal device in regulating student behaviour. Although it appears to serve as an instrument for recording student progress and so assist in providing evaluative feedback to students it was perceived by students as a symbol of authority and so served to implicitly regulate behaviour (inferred by their seriousness each time she monitored the actions of a particular group).

Distribution

In the site distribution was mostly used as a disciplinary tool. Distribution was used by the teacher to unify her actions and the content of lesson, as well as to organise students. By students, distribution mainly took the form of individuals differentiating themselves by actions such as coming in late, arranging themselves or their bodies in particular ways. Circles and group work were an interesting commonality across all Jenny's lesson sites.

Normalizing judgement and examination

Normalization occurred mostly in invisible ways in the incident, yet its prominence in visible ways was also noted: *Teacher: "grade 8's you know that we respect other people when they are presenting…"*. Invisibly, Jenny's body language ensured cooperation, at times it was her

presence either in the front of the class or close to a group that created a disciplinary effect. This was exemplified in cases where students looked in the direction of their teacher and stopped talking, smiling, fooling around etc. Normalising patterns were also implicit in student to student direction (his raised eyebrows were interpreted by his friend as a symbol of caution and he too became quiet and attentive…). Although the positions in space both of teacher and learners had served as explicit ways of normalising behaviour, their body language appeared to play a more dramatic role (Jenny's body language- stance, expression, gestures were used to gain attention and cooperation from learners).

In summary, the first layers of analysis revealed that: (1) these practices of power occurred in rapid succession, often overlapping, and were enacted by teachers and students alike--hence, supporting Foucault's view of power as circulating; (2) none of the sites was free of these techniques of power; (3) there are identifiable patterns in the circulation of power which enable claims to be made about the relationship between these techniques of power and various evaluation strategies such as peer-assessment, use of group-work, journals and portfolios.

From the snapshots of each technique at site, the following preliminary findings in terms of assessment can be reported:

- There was proportionally greater functioning of surveillance, regulation and distribution when an element of assessment was introduced into the setting by the teacher.
- There was a stronger functioning of normalisation in the setting when an element of assessment was introduced.
- The spread of techniques was greatest when the assessment element was introduced into the setting.

These points are opened for argument in the following section.

Issues and arguments

a) There was proportionally greater functioning of surveillance, regulation and distribution when an element of assessment was introduced into the setting by the teacher.

The teacher's strategy of 'mini circles' for the peer-assessment activity could be seen as progressive in the sense that it is a more humanistic, egalitarian arrangement of persons as opposed to the old "objective" row-by-row seating design of traditional assessment. This could also be perceived as 'in line' with the "inter-subjective" assessment philosophy of Outcomes-based assessment. Yet, ironically, from a Foucauldian standpoint, a circular arrangement is in fact a more effective panopticon, since every member of the circle is continuously visible to every other member, all the time. Again, even when one is in fact not being observed, at any moment one might be, and that is all that matters. Therefore all members of the groups (circles) in the classroom could be seen as conspirators in the panopticon, observers and observed; what the teacher did not see, others did.

Foucault argues that the panopticon, is not a simple physical structure, machine, or spatial arrangement: it becomes a way of life. As people accept the inevitability of being observed

and recorded, their habits change; they change. As people become more visible, the omnipresent circumstances that observe and record their lives become less visible. As the 'private' domain (the space of activity that is in principle unobservable, unrecordable) has become more and more circumscribed, an alarm about its now being 'invaded' seems ironic, for the real issue is with how that domain has become already so compromised; yet these restrictions, because they are often consensual, implemented gradually and with good sensible reasons, are actually more pervasive and insidious and hence harder to resist. This trend was evidently manifested in the patterns of resistance in the lesson: Resistance was evident on a very small scale and only on the entry of the "assessment" into the lesson. Resistance escalated with the introduction of the performance task, and proliferated during the first parts of the task and then died away completely. The resistance was a spontaneous reaction to something new, which when discouraged through the acts of surveillance by the teacher gave way to conformity, hence subjectification.

Foucault's work also highlights an interesting paradox: spaces of free action are increasingly circumscribed by restrictions on freedom (1977). These paradoxical tendencies come together in the idea of an 'architecture'—not merely the physical architecture of buildings (or panopticons), but the social architecture of ways of living. 'Architecture' can be seen as the locus where capabilities of creativity and mechanisms of control come together: architectures both contain and exclude, and the analysis of distinctive architectures can reveal important dimensions, and limitations, of human freedom. The 'group-work' context is seen as such an architecture in this study; while a medium that is enormously powerful and susceptible of immense collaborative learning (peer-assessment) learning uses, it includes constraints (as does any medium) on how information is shared, what sorts of information can be shared, and how people can communicate. It both enables and inhibits.

Within this educational contexts, this paradox is of fundamental importance for thinking about the relation of new technologies to learning and human freedom: the very same devices that allow the creation, exploration, and sharing of new knowledge and information, that spark new possibilities of action and interaction, also facilitate a heightened degree of observation and record-keeping about what people actually do.

One can avoid using such devices, in order to resist having one's freedoms compromised in one sense—but only at the expense of giving up the other kinds of freedoms and opportunities that those new technologies make available. A life without schooling—may in one sense be less panoptic, and more 'free'; but this freedom is obtained only at the cost of forgoing a number of other opportunities.

b) There was a stronger functioning of normalisation in the setting when an element of assessment was introduced

The relatively stronger functioning of normalisation than other techniques in the radical settings provides some explanation for the kind of 'resistance' educators encounter from students when they attempt to enact radical pedagogies. If power relations are inescapable in pedagogy, then whatever techniques are most strongly experienced are likely to encounter resistance.

Foucault said "there are no relations of power without resistances; the latter are all the more real and effective because they are formed right at the point where relations of power are exercised" (Foucault, 1980).

Together, these findings show, as Foucault puts it: *It seems to me that power is 'always already there', that one is never 'outside it'*, that there are no '*margins'* for those who break with the system to gambol in. But this does not entail the necessity of accepting an inescapable form of domination or an absolute privilege on the side of the law. To say that one can never be '*outside'* power does not mean that one is trapped and condemned to defeat no matter what. (Foucault, 1980, 141-2) It is also important, in all of this analysis, to remember that Foucault sees power as productive. For instance, he says of surveillance that it is a "mechanism inherent to pedagogy". In this analysis, my aim is not necessarily to suggest that we should attempt to alter power relations in these sites or that these configurations of power are negative. Instead, I am concerned, as a preliminary step, to attempt to understand in detail how power operates.

c) The spread of techniques was greatest when the assessment element was introduced into the setting.

The assessment activity, seen here as the 'examination', produced both positive and negative forms of energy for students and teacher alike. Students were clearly enjoying the task (evident from their body language and articulations) and this may be perceived as positive. On a negative note, the performance of the task made some students anxious and withdrawn (evident from body language). In a positive sense the teacher's plans were taking shape as students were co-operative, yet her constant anxiety with student control could be seen as negative. The teacher also appeared to reflect anxiety relating to the student perceptions of the task. It was almost as though she felt in-secure with the lack of formality surrounding the assessment and its formal implications in a formative sense: She constantly reminded that them that the assessment was important: "*Now guys remember…this is going into …your…um…portfolios hey….so do it properly*."

The introduction of the assessment into the lesson appeared to interfere with the order and formality that the teacher had established in her previous traditional seating arrangements. The teachers' regulative control increased and reached extreme proportions towards the end of the lesson. It was obvious that the teacher fore-grounded the 'controlling functions' of the assessment rather than its 'educational purposes.' The first of these are those reforms which focus on assessment, evaluation, and, more specifically, curriculum as the means to educational improvement. These contrast with reforms which aim at more independence at the classroom level, where teachers are given the freedom to match their instruction to their children. What she is pointing to is the fact that in many respects the efforts to control teachers and manage education serves instead to deaden instruction. By implication, the best hope we have to reach true excellence is to let go, to quit trying so hard to control. If we redirect the energy that we expend on these efforts to gain control, it might be applied to educational purposes instead of management ones.

Many reformers seem to feel that the only way to ensure that good education is going on in individual schools and classrooms is through good administration; supervision and accountability are the routes to good education (McNeil, 2002). What McNeil's analysis suggests stands in opposition to this 'common sense' point of view: good teaching can be chaotic and uncontrollable, and so the movement to control education directs us inevitably to bland mediocrity. The 'contradiction' is that efforts at control result in just the sort of defensive teaching that sticks as closely as possible to the defined curriculum and prepares students for the expected tests, which is what the reformers began by trying to change; the intellectual life is sucked out of the classroom.

In serving the social control function, the teachers, themselves both transmitting and being acted on by power, become part of the process by which the young are disciplined, and they themselves are controlled by the same forces (Foucault, 1977). The social control McNeil (2002) points to is two-edged; the students are controlled by the teachers, but both teachers and students are controlled and shaped in ways much more subtle and difficult to detect.

This is the working of power in its disciplinary form. Conformity is not the result of overt force that visibly bends the will of those subject to its operation; conformity results from the constant working of invisible constraints that bring us all toward the same 'normal' range of practices and beliefs (Foucault, 1977).

There are, we should note, serious deficiencies in Foucault's notion that everything reduces to power. His work at times becomes almost theological in tone; his faith that everything is reducible, finally, to power obscures the ordinary and valuable distinction between power and authority; his view that discipline is imposed on us as an effect of power makes us blind to Dewey's sense of discipline as a relationship between us and the world as we pursue our aims. Nonetheless, a Foucauldian analysis does serve to point to the extent to which we underestimate the complexity of 'the discipline problem' in schools, and the inadequacy of the pre-packaged programs sold to teachers as remedies.

Conclusion

In this paper, I have provided an overview of the study and of some of the findings that have emerged from the first site of coding categories applied to the data. The presentation is very technical and emphasises categorisation at the expense of contextualisation (Maxwell and Miller, 1993). From the outset I have wanted to conduct a systematic analysis of power relations in order to theorise from a strong empirical base. My aim, ultimately, is to produce a more flowing style that conveys in more narrative form the operation of power in these sites at the same time as I work towards a sociological theory of power relations in assessment.

References

- Bernstein (1990) Class, codes and control, vol. 4: the structuring of pedagogic discourse. London, Routledge.
- Bourdieu, P., & Passeron, J. (1977). *Reproduction in education, society, and culture*. Beverly Hills, CA: Sage.
- Clegg. S. R. 1989. *Frameworks of Power*. London: Sage.

- Department of Education (2000). Education for All: The South African Assessment Report. Pretoria: Department of Education.
- Foucault, M. (1977) Discipline and Punish: The birth of the prison. New York: Vintage.
- Foucault, M. (1980) *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, edited by Colin Gordon, Harvester, London.
- Gore, Jennifer M. 1996. 'Understanding Power Relations in Pedagogy'. Paper presented at the Annual Conference of the Australian Educational Research Association, Singapore.
- Hargreaves, A. (2002) "Perspectives on Classroom Assessment." In *American Educational Research Journal*, 39(1), 2002, 69-100.
- Mockler, N. and Sachs, J. (2002). 'A Crisis of Identity? Teacher Professional Identity and Evidence Based Practice'. Paper presented to the Annual Conference of the Australian Association for Research in Education. Brisbane, University of Queensland, December 2002.

‘Indigenous knowledge’, assessment, and international standards

Kai Horsthemke, University of the Witwatersrand School of Education

Abstract

‘African solutions to African problems’ has become a well- and widely-used slogan in the years since the transition to democracy in South Africa. More often than not, suggestions of such solutions are taken to constitute examples of ‘indigenous knowledge’. Conceptions of ‘local’ knowledge, development and struggles for cultural autonomy are usually articulated in terms of critiques of ‘western’ knowledge, development and hegemony. There are, broadly, two types of defence of ‘indigenous knowledge’. The more radical defence joins, and avails itself of, ‘post-colonial’, ‘post-modern’ and relativistic critiques of the ‘western’ notions of ‘universal knowledge’ and ‘international standards’. The more cautious approach emphasizes the different ways in which particular theoretical orientations—like realism, radical constructivism and anthropological cultural relativism—deal with questions of knowledge, including standards of assessment.

The present paper eschews the popular and ‘politically correct’ option of embracing the idea of indigenous knowledge. I argue here, on the basis of an analysis not of ‘indigeneity’ or ‘locality’ but of ‘knowledge’, that extant defences of this notion are deficient in several significant respects. The present analysis enables an understanding of the importance of context (of knowledge ascription and of assessment), without allowing the emphasis on relevance to erode any commitment to transcultural standards.

The idea of ‘indigenous knowledge’

The emphasis on ‘local’ knowledge is in part a response to the perceived lack of relevance and/ or sensitivity to social and cultural context manifest in standard assessment practices and reference to ‘global standards’. Thus, Malegapuru Makgoba tackles both the issue of (pursuit of) knowledge and the issue of standards, in university education in particular:

> The issue of pursuit of knowledge for its own sake and the so-called standards have always reflexly *(sic)* become contentious factors around the African university. … The pursuit of knowledge for its own sake has been one of the cornerstones of university education; but, is there such a thing as knowledge for its own sake today? Knowledge is a human construction that by definition has a human purpose. Knowledge cannot be sterile or neutral in its conception, formulation and development. Humans are not generally renowned for their neutrality or sterility. The generation and development of knowledge is thus contextual in nature. (Makgoba 1997: 177)

That knowledge ascription and justification have a crucial contextual component is surely not in doubt, but this does not mean that knowledge in all its aspects or components is essentially 'a human construction'. (I will return to these ideas below.) Nor does it mean that the pursuit of knowledge must be described and explained in consequentialist terms. It might be the *object* of knowledge that is and continues to be the legitimate cornerstone of university education. Makgoba continues:

> The second is the issue that Africanising institutions in context and curriculum or staffing them by Africans lowers standards. … The freedom of the African continent from colonial rule has one inherent dimension to it: the African must be free to do and choose as he deems fit. (Makgoba 1997: 177)

Whether one attaches an intrinsic or an instrumental value to education and knowledge, the deterioration in educational standards is evidently an absence of a significant good—and certainly perceived to be so by Makgoba, Barney Pityana and others. When Pityana speaks of the 'Africa-centred university' and of 'transformation as throughput rates and research profile', he intends this to be understood as a plea not for the lowering of standards but that South Africa focus on problems that are rooted in and significant for Africa (Pityana 2004: 4), including concerns around knowledge production and 'ownership'.

'Indigenous knowledge' is a fairly recent phenomenon, a buzz phrase that has gained increasing currency over the last quarter of a century, and in South Africa in particular since the first democratic election and subsequent political and educational transition in 1994. 'Indigenous knowledge systems' (IKS) is one of the National Research Foundation's focus areas (National Research Foundation 2005), and the introduction of IKS in school education is called for by the Revised National Curriculum Statements (RNCS) in the various learning areas (Department of Education 2002). Accounts or explanations of indigenous knowledge characteristically focus on 'indigenous'—as if this were the difficult or controversial term. It is commonly coupled with 'local', 'African', and the like, and contrapositioned with 'universal', 'global', 'world', 'western', 'Eurocentric' etc. (see Semali & Kincheloe, Higgs *et al*, and Odora Hoppers *passim*). To date, no policy document, statement or article has yielded the understanding of 'knowledge' the respective authors are working with—as if the concept of knowledge were simple and uncontroversial. I contend that it is not, and that a circumspect account of knowledge will reveal 'indigenous knowledge' to be something of a misnomer.

An analysis of 'knowledge'

'Curricula and classroom practices that accentuate multiple viewpoints and ways of knowing as well as the empowerment of students of color must be accompanied by assessments that acknowledge that there are many routes to similar outcomes, and that *knowledge is situational and culturally bounded*', according to A. Lin Goodwin and Maritza Macdonald (Goodwin & Macdonald 1997: 217; emphasis added). They illustrate their point by means of the following incident:

> Lena's classroom has been studying animals and the discussion has turned to zoos. For one of the assignments, children use drawings to depict their definitions of 'zoo'. One child, Kavemuii form Namibia, draws animals in a big, open, green, and gold field. The other children notice and begin to tease him, saying that he doesn't know anything about zoos. Lena notices and begins a discussion about how zoos came to be and why Kavemuii's drawing is so different from those of the other children. Through the discussion, the children come to understand that countries like the United States build zoos to house animals that are not indigenous to the country; zoos are designed to display animals in habitats that are made by humans and in areas that are inhabitable to animals. In Kavemuii's experience, animal habitats are very different; 'zoos' are the open plains where animals live free and not in cages. (Goodwin & Macdonald 1997: 217)

In concluding that this vignette 'illustrates how knowledge is personal, contextual, and cultural' (Goodwin & Macdonald 1997: 217), the authors arguably confuse two different kinds of knowledge. Indeed, they fail to acknowledge that there are, broadly, three kinds of knowledge. First, knowledge of a person, a place etc. is also referred to as familiarity- or acquaintance-type knowledge. Second, knowledge-how may be called practical knowledge, even skill. Third, knowledge-that constitutes propositional or declarative (often also referred to as factual or theoretical) knowledge. The incident related by Goodwin and Macdonald appears to involve familiarity- or acquaintance-type knowledge, in which case their conclusion is plausible. To the extent that it also involves knowledge-that, however, the conclusion that 'knowledge is personal, contextual, and cultural' is not only unwarranted but is clearly avoided with the help of precise definition, namely of 'zoo'.

Whereas the first two types of knowledge are fairly straightforward, it is the third type of knowledge that is more complex. Traditionally defined as 'justified true belief', knowledge-that has three essential components:

- (strong) belief
- truth
- and (suitable) justification.

Each of these is necessary, and jointly they are sufficient, for knowledge. In other words, if one were missing, it would be meaningless to speak of 'knowledge'. 'Belief' is its subjective component, while 'truth' constitutes the objective anchor for knowledge. While beliefs may vary from individual to individual, society to society, culture to culture—and indeed in terms of strength and duration—truth does not so vary. Truth refers to what is the case, independently of what individuals believe, think or feel may be the case, independently even of public or general consensus. The third component, 'justification', has a kind of bridging

role between the subjective and the objective, between belief and truth. The justification component is multifaceted. Thus, what counts as *suitable* justification is determined by

- degree
- kind
- and context of justification.

Degree and kind of justification, and 'authentic assessment'

As far as the requisite *degree* of justification is concerned: minimal justification is clearly not enough, while conclusive justification is usually not available. Normally, that is, other than in mathematics and deductive logic, we accept justification that is less than conclusive, that is, reasons that are nonetheless compelling. Different *kinds* of justification include observation, sense experience, introspection, memory, oral and written testimony, deductive and inductive reasoning, and so on. The present account of suitable justification bears on the attribution of knowledge insofar as it concerns both self-ascription and other-ascription, the framing question being: 'Under what conditions can I/ others be said to know?' There is a parallel here between knowledge attribution and both self- and other-assessment. 'Authentic assessments', according to Goodwin and Macdonald,

> are often described as meaningful and comprehensive measures of what learners know and are able to do. … [They] are characterized by continuous observations of learning, depth and breadth of response, cycles of revision and refinement, students' engagement in self-assessment, and connections between what is being assessed and real-world issues and questions. … Authentic assessment begins with teachers making it their business to purposefully watch, listen to, talk with, and think about the children in their classrooms. By observing, recording, informally monitoring, conferencing with, and interviewing their students, teachers initiate an ongoing process that uncovers who learners are and what they know, and that leads to opportunities for teachers and children to build shared language, meaning, and beliefs. (Goodwin & Macdonald 1997: 211, 223)

These sources of what Goodwin and Macdonald refer to as 'holistic assessments that … enable children to demonstrate learning by integrating and applying knowledge and skills to real-world tasks' (Goodwin & Macdonald 1997: 223 fn.1) are similar to sources of justification, that is, on the basis of which educators ascribe knowledge to their learners and students—and, indeed, on the basis of which we ourselves can claim to know.

Knowledge, justification and the 'social'

Stewart Cohen's verdict that the 'social component [of knowledge] is best seen as indicating that attributions of knowledge are context-sensitive' (Cohen 1986) is related to what Israel Scheffler says about the suitability of justification[1]. Scheffler argues that the idea of *suitability* 'involves standards, which are normally applied more strictly in some cases, more approximately in others, thus giving rise to multiple interpretations of *knowing*' (Scheffler 1965: 96). Thus, the justification component permits some kind of leeway. What counts as

[1] Scheffler initially refers to 'evidential adequacy', a notion he later abandons in favour of that of 'the right to be sure'. I take 'justification' to be more comprehensive than 'evidence' and to be consistent with the notion of the right to be sure.

suitable justification in the case of a young child or learner or person from a remote rural area, with limited opportunities, resources or access to information, differs from that required of an older, more mature child or learner or person from an industrialized, technologically advanced/ privileged, urban background. Yet, in all the various cases, the justified belief must be true. In the absence of truth, one cannot meaningfully speak of, or ascribe, knowledge. Scheffler suggests a subtle shift from examining beliefs to examining the *contexts* in which beliefs are advanced as knowledge-claims. In other words, he suggests that we distinguish the question concerning justification (of a belief) from the 'question of *appraisal of the believer*' (Scheffler 1965: 102). 'To speak of the right to be sure is, in the present context, to appraise the *credentials* of belief from the vantage point of our own standards; it is to spell out the attitude of these standards toward specific *credentials* offered for a belief', Scheffler contends[2] (Scheffler 1965: 102).

Like Scheffler, Cohen argues that justification, or having good reasons, is relative to an epistemic community. He advances his argument through an analysis of what it means to have good reasons for believing something. The concept of *defeasibility* is crucial to Cohen's argument. One's reasons for believing something are *defeasible* if there is something else that could count against them, that is, something that could *defeat* them or undermine their feasibility. According to Cohen, we can say that someone (like a six-year-old) has good reasons if, given her reasoning ability, it is (epistemically) permissible for her to believe that something is the case. In Scheffler's words, we would be inclined to apply our standards of justification more leniently in the case of the six-year-old and more strictly in the case of the sixteen-year-old.

The important point for educators is that what counts as a good reason depends on who is giving the reason and in what context. One of the responsibilities of an educator is to assess learners' knowledge in a way that is sensitive both to their level of understanding and to the context of assessment. Another (related) responsibility is to develop learners' grasp of the intersubjective standards of different learning areas – or, in Cohen's terms, to help learners move from a level of reasoning which provides *subjectively evident* grounds for believing something to a level which provides *intersubjectively evident* grounds for belief.

The concept of having good reasons is not without ambiguity. A person can have *subjectively good reasons* (that is, reasons that are clear and convincing to her, given her level of understanding) or *intersubjectively good reasons* (that is, reasons that are clear to her and that comply with the standards of reasoning of the social group to which she belongs). How are these different applications of justification relevant to the concept of knowledge? When is an educator entitled to say that a learner knows something (in the sense of *knowing that*)? To put the question more formally: Under what conditions may an educator attribute knowledge to a learner? According to Scheffler, when we judge that someone possesses suitable justification, we are judging that he possesses reasons the quality of

> which he understands. In saying he knows, we are not merely ascribing true belief but asserting that he has proper credentials for such belief, the *force* of which he himself *appreciates*. (Scheffler 1965: 112)

[2] The notion of 'the right to be sure' is A.J. Ayer's (Ayer 1956).

Scheffler's and Cohen's arguments imply that even if a learner has subjectively good reasons for believing something to be true, she does not have knowledge unless she also has *intersubjectively good reasons* (and, of course, unless her belief is true). One of the tasks of an educator will be to assist learners to acquire the relevant concepts and intersubjective standards of justification. To repeat, I am not talking about *perfect* reasons: I am talking about *good* reasons. An important feature of what constitutes *good* reasons is that they are *reliably produced*. If reasons are unreliably produced, they cannot function as justification for one's belief/s. What does 'reliably produced' mean? For one thing, one's sense experiences must be reliably connected with the world, one's sense organs must be intact, etc. For another, one's reasoning must be correct. This analysis of good reasons indicates why reference to them is context-sensitive. Neither our reasoning nor our sense-experiences are infallible. Nonetheless, if they are generally reliable sources of justification, the reasons they produce might be called *intersubjectively certain*. As Cohen says: 'Reasons can be permissible grounds of belief, relative to that standard, even though they are not ideally correct' (Cohen 1986: 575).

If reference to good reasons is context-sensitive, does this mean that the criteria for knowledge-ascription change with the respective social group? Is knowledge itself relative? In Plato's cave parable, whatever the enlightened person knows about 'reality' stands in stark contrast to the (majority) view that what the prisoners in the cave claim to know is reality. Does this indicate that *knowledge* is ambiguous between various concepts, each based on a different standard? Is this knowledge context-dependent? Scheffler's and Cohen's arguments suggest that it may be better to say that attributions of knowledge are *context-sensitive*. This is because the term '*context-sensitive*' does not offer an open invitation to epistemological relativism.

Epistemological relativism and relativism about truth

What are epistemological relativism and relativism about truth? In a nutshell, relativism about knowledge and truth denies that there can be any objective knowledge or truth that is not dependent on some specific social or historical context or conceptual framework. Ernst von Glasersfeld, in defending what he has called 'radical constructivism', asserts:

> There is no simple argument to justify the distinction between experiential reality and ontic reality. … As a constructivist, I have never said (nor would I ever say), that there is no ontic world, but I keep saying that we cannot know it. I am in agreement with Maturana when he says: 'an observer has no operational basis to make any statements or claim about objects, entities or relations as if they existed independently of what he or she does'. (Von Glasersfeld 1991: § 17; see Maturana 1988: 30)

He says that he, too,

> arrived at this conclusion, albeit by a path quite different from [Maturana's]: I started from the sceptics, he from biology. The crucial point is that we do not make claims of knowing what exists 'in itself', that is, without an observer or experiencer. … And as far as our knowledge … is concerned, I claim that we cannot even imagine what the word 'to exist' might mean in an ontological context, because we cannot conceive of 'being' without the notions of space and time, and these two notions are among the first of our conceptual constructs. (Von Glasersfeld 1991: § 18)

Willard Van Orman Quine and J.S. Ullian concur that an 'observation is made by an individual'. However,

> the truth of the observation sentence is an intersubjective matter. Here a favorite old irrationalist doctrine finds both its seductiveness and its rebuttal. The hoary view contends that truth is relative to believer; there's truth for me and truth for you, and their reconciliation is generally neither possible nor desirable. Now the variable ownership of acts of observation might be cited in support of this doctrine …; but which observation sentences are *true* will not … depend on either of us nor on any other observer. (Quine & Ullian 1978: 28)

The 'hoary view' or 'irrationalist doctrine' Quine and Ullian are taking on in the passage from which the quote is lifted is relativism about truth. To be a relativist about truth is to maintain that there is no universal, transcultural or objective truth, that 'truth' is in the eye of the beholder, or in the mind of the believer: it differs from individual to individual, from society to society, from culture to culture. In other words, truth is particular or relative to a specific personal, social, cultural, historical or geographic context.

The defence of relativism draws on the central role that Quine and Ullian, and others, have accorded to observation, with regard to belief systems and knowledge formation. Observations are made by individuals, and 'the ultimate evidence that our whole system of belief has to answer up to consists strictly of our own direct observations – including our observations of our own notes and of other people's reports' (Quine & Ullian 1978: 21). Given, furthermore, that our observations differ, does this not imply that the truth/s that we access by means of our observations will differ? No, say Quine and Ullian: our observations may indeed differ—the may either contradict each other, or they may be of different aspects of the same object—but what is actually the case, what is true, does not depend on these observations, nor on any observers.

'Truth is relative to believer'. Is this truth (if it is that) also relative to believer? If so, why should it impress others? If not, then there exists at least one truth that is not relative. 'There's truth for me and truth for you, and their reconciliation is neither possible nor desirable'. Is this my truth? Or is it also your truth? If the former, why should it impress those who hold a different view of truth? If the latter, this indicates that reconciliation is possible—yet, again at the expense of relativism and in favour of universalism. Either way, the relativist will be caught up in paradox, in a logical conundrum. At some point, he will want to claim that his statements about the relativity of truth are, in fact, universally (that is, nonrelatively) true—which he cannot do consistently, given his relativism. If anything, von Glasersfeld's characterization of 'scientific truth', as requiring the 'separation of metaphysical beliefs and convictions, which purport to reflect an ontological reality, from rational/ scientific knowledge, which is given an instrumental function in the living organisms' management of their subjective experiential reality' (Von Glasersfeld 1991: § 48), achieves the opposite of what it is intended to do. It provides no means for distinguishing between scientific truth and superstition.

Moreover, if 'an observer has no operational basis to make any statements or claim about objects, entities or relations as if they existed independently of what he or she does' (Von

Glasersfeld 1991: § 17; see Maturana 1988: 30), how can the observer *claim* that these objects' existence *depends* on what s/he does? How does von Glasersfeld *know* that we *cannot* know the ontic world? What he refers to as the 'original seed of constructivist ideas', 'undoubtedly the sceptics' realization that we can have no certain knowledge of the real world' (Von Glasersfeld 1991: § 7), is problematic in that sceptics cannot claim to 'know' or to have 'realized' this. Similarly, the 'sceptics' irrefutable proposition that the truth of what we would call "knowledge of the world" cannot be assessed or demonstrated because the "representations" of which it is supposed to consist can never be compared with what they are supposed to represent' (Von Glasersfeld 1991: § 48) cannot, for reasons of consistency, be taken as 'certain knowledge of the real world'. Nor can it be taken as reflecting 'experiential reality', given that 'the sceptics [have] demonstrated quite irrefutably that the senses are fallible' (Von Glasersfeld 2003). It follows that von Glasersfeld and the sceptics cannot both construct their cake and eat it.

How is this to be applied to the discussion of indigenous knowledge? There may be a sense in which knowledge of the familiarity or acquaintance type can be 'indigenous'. 'Indigenous' South Africans could be said to have such knowledge of their traditions, customs, myths and folklore, and the San have such knowledge with regard to the terrain in which they live, that is uniquely and distinctly theirs. Similarly, the notion of 'indigenous practices' or 'indigenous skills' appears to make sense—like pottery or weaving practices, or architectural and design skills. But does the idea of 'indigenous' knowledge make sense when applied to the third type, factual or theoretical knowledge? I would suggest that everything depends on the truth condition. The idea of 'indigenous beliefs' certainly is plausible, and there may well be some leeway with regard to justification, to enable greater leniency in assessing (the justification for) Ju'hoansi knowledge claims and greater stringency in assessing those of university graduates. But could there be 'indigenous truths', could there be (an) African truth(s) that is/ are different from 'western' or 'occidental' truth(s)? If an affirmative answer is given to these questions, there are two problems. The first is that of superstition. Anything that is honestly and (more or less) justifiably believed and advanced as a knowledge claim could count as truth. There would be no way to distinguish between knowledge and superstition. The second problem is that of relativism. If knowledge and truth did differ from individual to individual, society to society, culture to culture, then it would be impossible to pass judgment on another's knowledge claim. Moreover, in order to decide what is true or false, one would merely have to consult the beliefs prevalent in one's own society. Finally—if one thinks especially about standards over time, across subjects and across learning pathways—one could not really say whether any progress has been in a society, in terms of advancement in knowledge and assessment. (Think of the views of a 'flat-earth' society in all these instances.) The deeper problem is that relativism about knowledge and truth is expressible only as knowledge claims or as stating a truth—which would, at best, not be very compelling or, at worst, incoherent.

Concluding remarks

To sum up, there are several questions one might ask, that bear also on the plausibility of applying the formula 'everything is relative' in discussions about assessment and educational standards, with regard to the 'hoary' conception that concerns the status of truth. *Is* truth (for example, facts about the shape of the earth) relative? *Does* it differ from culture to culture, from society to society, and even from individual to individual (as in 'What is true

for you may not be true for me')? Or is what may differ from culture to culture etc. not truth but *beliefs*? At this juncture, one may stop to think about why beliefs are referred to as the 'subjective' component of knowledge and truth as the 'objective' component. Can one compare and evaluate different beliefs and belief-systems or not? Do we have recourse to an objective framework of reference? Finally, is it *true* that truth is relative? It is worth also noting the implications of this for the status of *knowledge*. Truth, after all, is a necessary condition for knowledge, as are belief and justification. Philosophically, therefore, does it make sense to speak of *true* knowledge or *legitimate* (or *valid*) knowledge? Can there be any *other* kind, knowledge other than 'true' and 'legitimate' or 'valid'? Does it make sense to say that knowledge is 'relative'? Or is what may differ from culture to culture, from society to society, from individual to individual, again, either beliefs *or* the level (or standard) of justification required or expected? Clearly, the trick for the epistemological relativist is to avoid advancing his thesis as anything other than a nonrelative knowledge claim. If he presents his epistemological relativism as a piece of universal knowledge, he can be accused of inconsistency. Yet, if he does not, it is unlikely to convince nonrelativists—just as realists would be unfazed by the construction of radical constructivism.

By contrast, the term *context-sensitivity* indicates that the standards of knowledge-attribution and of assessment may be determined by (1) the context of attribution/ assessment and (2) the *intentions* of those who attribute or ascribe knowledge and who are involved in assessment. It does not imply that truth, and with it (the concept of) knowledge itself, is relative to a particular standard, unstable, or changing.

Where does all this leave the notion of indigenous knowledge? If the important term here is 'indigenous', then it refers either to indigenous practices or to indigenous belief. It would consequently be wise to jettison the term 'knowledge'. On the other hand, if it actually is meant to refer to 'knowledge' in the factual or propositional sense, then the idea of 'indigenous' knowledge simply fails to make sense. What we are dealing with here is knowledge as such.

In South Africa as elsewhere, people do not all have the same cognitive resources, skills and opportunities; they do not act or operate free of time constraints. Their situations are characterised by different levels of expertise, by different opportunities for information gathering, by different levels of cognitive maturity and training, and by severe time constraints. As Alvin Goldman cautions, a 'social epistemology for the real world needs to take these constraints into account' (Goldman 1992: 223). If the present analysis is plausible, it arguably enables an understanding of the importance of context, in terms of both attribution and provision of knowledge, as well as of assessment, without allowing the emphasis on relevance to erode any commitment to transcultural standards.

References

- Ayer, A.J. 1956. *The problem of knowledge*. Harmondsworth: Pelican
- Cohen, S. 1986. Knowledge and context. *Journal of Philosophy* 83: 574-586
- Department of Education. 2002. Revised national curriculum statement Grades R-9 (Schools). http://www.education.gov.za/content/documents/559.pdf (accessed 8 April 2005)

- Goldman, A.I. 1992. Liaisons: Philosophy meets the cognitive and social sciences. Cambridge/ Massachusetts: MIT Press
- Goodwin, A.L. and Macdonald, M.B. 1997. Educating the rainbow: Authentic assessment and authentic practice for diverse classrooms. In Goodwin, A.L. (ed.) 1997. *Assessment for equity and inclusion: Embracing all our children*. New York & London: Routledge: 211-227
- Higgs, P., Vakalisa, N.C.G., Mda, T.V. and Assie-Lumumba, N.T. (eds.) 2000. *African voices in education*. Lansdowne: Juta
- Higgs, P., Lillejord, S., Mkabela, Q., Waghid, Y. and Le Grange, L. (eds.) 2004. Indigenous African knowledge systems and higher education. *South African Journal of Higher Education* 18 (3)
- Makgoba, M.W. 1997. Mokoko: The Makgoba affair. A reflection on transformation. Florida: Vivlia
- Maturana, H. 1988. Reality: The search for objectivity or the search for a compelling argument. *The Irish Journal of Psychology* 9 (11): 25-82
- National Research Foundation (NRF) 2005. Guide 2006: The NRF Focus Area Programme. http://www.nrf.ac.za/focusareas/iks/ (accessed 10 March)
- Odora Hoppers, C.A (ed.). 2002. Indigenous knowledge and the integration of knowledge systems: Towards a philosophy of articulation. Claremont: New Africa Books
- Odora Hoppers, C. 2002. Indigenous knowledge and the integration of knowledge systems: Towards a conceptual and methodological framework. In Odora Hoppers, C. (ed.) *Indigenous knowledge and the integration of knowledge systems: Towards a philosophy of articulation*. Claremont: New Africa Books: 2-22
- Odora Hoppers, C.A. 2005. *Culture, indigenous knowledge and development: The role of the university*. Johannesburg: Centre for Education Policy Development (CEPD), Occasional Paper No. 5: 1-50
- Pityana, B. 2004. If not now, then when? *Beyond Matric/ Mail & Guardian*, May 7-13: 1,4
- Quine, W.V. & J.S. Ullian. 1978. *The Web of Belief*, revised 2nd edition. New York: McGraw-Hill
- Scheffler, I. 1965. Conditions of knowledge: An introduction to epistemology and education. Illinois: Scott, Foreman & Co.
- Semali, L.M. and Kincheloe, J.L. (eds.) 1999. *What is indigenous knowledge?: Voices from the academy*. New York & London: Falmer Press
- Semali, L.M. and Kincheloe, J.L. 1999. Introduction: What is indigenous knowledge and why should we study it? In Semali, L.M. and Kincheloe, J.L. (eds.) *What is indigenous knowledge?: Voices from the academy*. New York & London: Falmer Press: 3-57
- Von Glasersfeld, E. 1991. Knowing without metaphysics: Aspects of the radical constructivist position. In Steier, F. (ed.) *Research and reflexivity: Inquiries into social construction*. London: Sage Publications: §§ 1-55
- Von Glasersfeld, E. 1997/ 2003. Homage to Jean Piaget (1896-1982). *The Irish Journal of Psychology* 18 (3): 293-306 (http://www.umass.edu/srri/vonGlasersfeld/onlinePapers) (accessed May 2005)

Ethnobased learning and international standards in assessment

Toliwe Chehore, Zimbabwe School Examinations Council

Abstract

The local environment in which people are living is rapidly changing and at the same time, being increasingly affected by and integrated into the large global environment. Those affected include skills, values and competencies required to live a sustainable life. Education is now required to impart learning both local and transnational knowledge. Assessment should be reformed from that which assesses the traditional fact-based knowledge to assessing that which reflects 'real world' experiences and requires students to structure the task, apply information, construct responses and explain the processes by which they arrive at answers. Measuring instruments must consist of meaningful tasks that demand problem-solving, critical thinking, and good writing skills.

In this study the November 2005 Mathematics, Computer and Science subjects were analyzed in terms the context of the question and skills tested. Although localization of examinations in Zimbabwe changed the context of the content to focus on local knowledge, the teaching, learning and assessment strategies remained unchanged resulting in a misalignment within the education system. This could have been caused by trying to maintain internationally competitive standards. The misalignment can result in Zimbabwe lacking behind in current trends in educational reform and as such educational standards no longer being internationally competitive.

Introduction

It is the desire of every parent and stakeholder that the learner receives quality education of high academic standards. Academic standards are considered high if the knowledge and skills acquired at schools are used in everyday life. Hence most countries are searching for ways on how to educate all their learners to high standards which are internationally competitive. Internationally competitive standards are being sought after, because nations can no longer stand apart from each other socially, politically and economically. There is increased

interdependence of national economies and increased economic competitiveness. These come together as a global market where human capital and quality of education are matters of prime importance. Education is all about acquisition of knowledge. Current trends in educational theories emphasize that learning occurs best when knowledge is derived from the learners' experiences. The purpose of this paper was to link teaching of local knowledge with internationally competitive standards of education through ethno-based learning.

Background

Globalization

People are now living in a rapidly changing environment with vast and complex sources of information. The movement of information, together with that of people, goods, and ideas is always accelerating across the globe such that although human lives continue to be lived in local environments, these are continually being integrated into the larger global world. This has an impact on education systems worldwide, because education has to prepare learners for the costs, challenges and opportunities of both the local and international environment. The greatest challenge to the learner is how to manage change and difference. People in today's world interact with others of different racial, ethnic, national, linguistic and religious backgrounds. They are then required to understand human nature, history, different cultures, values, traditions and social organizations. This requires individuals who approach and solve problems from multiple perspectives, and are able to work collaboratively in groups made up of diverse individuals. Globalization requires a paradigm shift in education from a focus on mastery and regurgitation of rules, facts and principles to the development of both non-technical and technical abilities required in real life. As such current trends in educational reform reject the acquisition of pure knowledge without any application to real life situations. Life long learning, higher order cognitive skills, intra and interpersonal skills are required in this ever changing world. Higher order cognitive skills are required for multidisciplinary understanding of society, problems and solutions. Intrapersonal skills are needed for personal understanding and development and interpersonal skills for better human relationships. The goal of education should be that each and every child should be equipped with skills that are required of an individual to function after school. According to Cotton (2001) the fundamental skills possessed by an educated person include the ability to think carefully, reflectively, critically, creatively, make wise choices and right decisions. Some of the skills required of school graduates in terms of business, employment and life in general are:

1. higher order skills such as
    - application, analysis, synthesis, and evaluation of concepts in real life contexts
    - creative and innovative thinking
    - critical thinking
    - problem solving
    - logical, deductive and inductive thinking
    - decontextualisation
    - transfer of knowledge between disciplines, issues and subjects
2. metacognition
    - self-confidence

- positive self-image
- self-discipline
- self-management

3. intrapersonal skills
    - independence
    - responsibility
    - honesty
    - enthusiasm and motivation
    - dependability
    - integrity
4. interpersonal skills
    - cooperation and team membership
5. basic skills and oral communication (speaking and listening), including
    - understanding and following instructions, mathematical computation
    - and simple writing.

A lack of the majority of these has been found to be some of the causes of improper work habits behavior, attitude and a general failure in life. Cotton's research (2001) showed that these capacities were not genetic but are deposited qualities that are possessed by the owner, but could be developed and through instruction and practice from school level. The content of what is taught comes from the curriculum and is assessed during and after instruction. An education system should have its curriculum, instructional strategy and assessment aligned to achieve desired goals.

The content

Generally, the curriculum contains what learners should know and be able to do, (content), how it is taught (instruction) and how it is measured (assessment). Due to globalization and availability of vast sources of information the content should include:

- different ways of knowing and validating information
- approaches from multiple perspectives
- identifying and connecting ideas, concepts and applications
- communication
- metacognitive abilities
- awareness of contributions of different cultures to current knowledge
- context that is relevant to the real world environment.

The curriculum should be suitable for all learners irrespective of the background and give all learners the opportunity to learn. It should promote the acquisition of local, regional and international knowledge.

Instruction

Effective instruction cannot be prescribed for effective teachers because it involves continuous adjustment between the teacher and the learner. But generally instructional strategies should:

- be meaningful and motivating
- inspire learners to learn more and with enthusiasm
- be relevant to their lives, including in pursuit of work and leisure
- adequately prepare learners for the society in which they live and work
- be based on real world experiences
- should use prior knowledge as a base for new knowledge.

Ethno-based learning and local knowledge

Ethno-based learning, a concept derived from ethno-mathematics, is an instructional approach based on the premise that learning occurs best when it is derived from the real life experiences of the learner. Every learner has knowledge of one's local environment and this, together with one's experiences constitute local knowledge. According to Pandey (1990) learners construct new knowledge by integrating it with prior knowledge. This knowledge has meaning and value because it relates directly to real life experiences, All learners believe that they can learn and are keen to learn new concepts. Ethno-based learning encompasses the fundamental theories of constructivism, contextual teaching, and problem-based learning. Constructivism, according Doolittle (1997) recommends that:

- learning take place in authentic and real world environments
- content and skills be made relevant to the learner
- content and skills be understood within the framework of the learner's prior knowledge
- learners be encouraged to be self-regulatory, and self-aware
- teachers provide for and encourage multiple perspectives of content.

Wilson (2001) describes "contextual teaching as a concept that involves connecting the content that the students are learning with the context in which that content could be used". Howey, (1998) in Wilson, (2001) notes that contextual teaching and learning emphasizes acquisition of higher-order thinking skills and knowledge transfer:

> Problem-based learning is an instructional approach that uses real world problems as context for students to learn critical thinking and problem solving skills and to acquire essential concepts of a course. (Wilson, 2001).

Authentic assessment

If educational context, content, and instructional practices are being reformed, then the assessment and evaluation strategies have to be changed and aligned with the new approaches. Ethno-based learning activities should be aligned with not only the instructional objectives but with the subsequent assessment. The use of the traditional pen and paper assessments only, where ethno-based learning has taken place, causes misalignment of what is expected of learners. The primary characteristic of ethno-based learning is that learning is anchored in the real world of the learner. Authentic assessment tasks together with traditional standardized assessment tasks should therefore be used to assess learning achievement. An authentic task is an assignment s designed to assess the ability to apply knowledge and skills to real world challenges. Ethno-based learning can therefore be aligned with authentic assessment. This is because what is important is learners use knowledge

acquired, in the face of real life challenges, rather than how much knowledge they would have acquired.

Ethno-based learning and authentic assessment produce learners who are:

- highly motivated and committed to learning
- equipped with a range of transferable skills
- are equipped with more meaningful and relevant knowledge that goes beyond the classroom
- collaborative and cooperative
- deep learners rather than surface learners
- life long learners
- reflective, analytical, inquisitive, critical, innovative and creative
- metacognitive
- fairly and reliably assessed.

(Adapted from Mowl, G. (1996) *Innovative Assessment*)

Assessment standards

These are determined by the needs of the society. They should be set such that they measure the acquisition of qualities required of individuals to live a sustainable life. Due to globalization, standards have to be tailored towards both the local and international communities. Quality assessment standards are based on principles of assessment namely, validity, reliability, and fairness. The assessment standards should promote acquisition of local and international knowledge and teaching practices that enhance learning by all children, irrespective of their background and abilities.

Localization of examinations in Zimbabwe

Secondary school examinations in Zimbabwe, were, before independence, set and marked by the University of Cambridge Local Examinations Syndicate (UCLES). Soon after independence, Zimbabwe embarked on a process of localizing the whole education system, examinations included. This meant that the curriculum was developed and examinations marked and processed locally. The content of the curriculum and the context of the examinations were to be locally based. The teaching strategies, examination techniques, grading systems and certification remained unchanged. The standards therefore remained unchanged and international as the graduates were and are still being accepted in many countries across the world. The history of education in most countries, Zimbabwe included, shows that initially secondary education was a preserve for a minority and the curriculum was dominated by concepts which prepared learners for university education. Mass secondary education, which is now prevalent in most countries like Zimbabwe, requires that learners be prepared not only for university education, but for work, citizenship and membership of both the local and international community. The challenge was how to educate all learners, irrespective of background to international standards. This research was important in that it highlighted the need to align instructional and assessment strategies of Zimbabwe's education system, in line with educational current trends, if its products are to remain internationally competitive.

Aim

The aim of this paper was to link, using the experience of the Zimbabwe School Examinations Council (ZIMSEC), teaching of local knowledge, through ethno-based learning and authentic assessment, to the achievement of international standards of education. The objectives of the research were to determine the extent to which questions set in Ordinary level (O'Level) mathematics and science subjects

1. were in context of out-of-school experiences
2. tested abstract concepts, higher order skills and non-cognitive skills.

Methodology

Seven most popular mathematics and science subjects at O' level were selected because more than seventy five percent of secondary school children leave school after this level, having sat for at least, two of the subjects. A total of twenty three papers were analyzed from the November 2005 examinations. Each question in each component of the subject was analyzed terms of its context and skills tested. Marks allocated to the question or parts thereof were summed up per paper so as to determine the proportion of different contexts and skills involved in the question. Assuming that the papers were sat according to the marking scheme, the percentages of the higher and lower order skills were obtained, together with the paper weighting, from the assessment schemes in the syllabuses. All the science syllabuses had three categories of assessment objectives, namely:

1. knowledge and understanding
2. handling information and problem solving
3. experimental practical skills.

The first two objectives were tested in theory papers and the third in the practical or alternative to practical papers. The first objective was interpreted to test lower order, whilst the other two tested higher order skills.

Results

Table 1.1 Skills and Concepts tested

Subject	Paper	Paper weighting	Higher order skills	Lower order skills	Abstract concepts	Out of school experience	School experience	Total
		%	%	%	%	%	%	Marks
Integrated Science	1	30	30	70	0	38	62	40
Integrated Science	2	50	30	70	0	12	88	100
Integrated Science	3	20	100	0	0	42	58	40
Biology	1	30	45	55	0	58	42	40
Biology	2	50	45	55	0	44	56	100
Biology	3	20	100	0	0	100	0	40
Biology	4	20	100	0	0	75	25	40
Physical	1	30	45	55	22	10	68	40

Science								
Physical Science	2	50	45	55	10	48	42	100
Physical Science	3	20	100	0	0	50	50	40
Physical Science	4	20	100	0	0	0	100	40
Chemistry	1	30	45	55	40	0	60	40
Chemistry	2	50	45	55	18	7	75	100
Chemistry	3	20	100	0	0	0	100	40
Chemistry	4	20	100	0	0	0	100	40
Physics	1	30	45	55	20	45	35	40
Physics	2	50	45	55	5	40	55	100
Physics	3	20	100	0	0	0	100	40
Physics	4	20	100	0	0	0	100	40
H. Social Biology	1	70	45	55	7	18	75	100
H. Social Biology	2	30	45	55	10	40	50	40
Mathematics	1	50	-	-	34	16	40	100
Mathematics	2	50	-	-	27	30	43	100

All papers ones, except for Mathematics but including Human and Social Biology paper two and not paper one, were composed of forty multiple choice questions. These papers had a weighting of thirty percent and the ratio of marks for higher to lower order skills was forty-five to fifty-five percent, except for Integrated Science which had thirty to seventy percent. The range of marks for out-of-school experiences was zero to fifty-eight percent with Biology only, having more than fifty percent.

The paper twos, for the same category of subjects, were composed of sections of short-structured and free response questions. They had a weighting of fifty percent and a mark total of one hundred. Performance in paper twos alone, in all science subjects and paper one in Human and Social Biology had a much greater effect on the final grades than any of the other papers in the same subject. These papers had fifty-five percent of the marks in lower order category of skills, except for Integrated Science which had seventy. The range of marks for those questions which in context of-out- of school experiences was from seven to forty-eight percent.

The practical papers which had out-of-school experiences ranging between zero and one hundred percent were all, by their nature, of higher order skills.

Mathematics had a range of assessment objectives which could not be classified the same way as the science subjects. These were, at least, from the level of application. Mathematics together with, physics and chemistry had marks ranging from five to forty percent for questions testing abstract concepts.

Theory papers alone accounted for at least eighty percent of the marks, and in almost all cases the subject had an alternative to practical paper. This meant that all papers could be assessed through pen and paper alone. None of the papers had a school-based or continuous assessment component, that is one based on assessment over a prolonged period.

Except for Biology papers, questions based on abstract concepts plus in-school experiences were in the majority in all subjects especially in Mathematics. All subjects had theory papers whose total weighting was at least eighty percent, and these were dominated by questions testing lower order skills, as per assessment scheme.

None of the questions tested non-cognitive and metacognitive skills.

These were standardized tests from a syllabus whose content was local but taught using the traditional fact-based teacher dominated approach. They sought to discriminate among different learners and only a minority achieved proficiency in the different disciplines. The graduate is unlikely to be of the calibre currently required by the global society.

Discussion

It can be concluded from the background of the study, that the content, context, teaching, learning and assessment standards are all driven by the needs of the society. These needs although locally based are influenced by the effects of globalization. This can be summarized in the form of a model, (see Appendix), linking educational components of content, resources, instruction and assessment. The required quality, hence standard, of these, are determined by the characteristics required of a school graduate. These in turn are determined by the needs of the society, which today are largely influenced by globalization. The characteristics required of an individual today are at the center, as they should be taken into account at each and every stage of the education system. All the components of the education system should be aligned to produce the desired products.

The Zimbabwe education system was misaligned by localization. Although the content was locally based, all the other components remained unchanged. As such the qualities of the resultant graduate were largely unchanged. This is now causing misalignment with the needs of the society, which requires its members to fit into the global society.

Conclusion

Alignment of the education system components removes tension between teaching of local knowledge and internationally competitive standards of education. Local knowledge can be used both as a resource and a strategy whilst standards are used to indicate certain qualities required in the product of a process.

Recommendations

The Zimbabwe school education system should have its components aligned and reformed along with current global trends of teaching, learning and assessment. Ethno-based learning and authentic assessment should complement localization of the content and context of education in Zimbabwe.

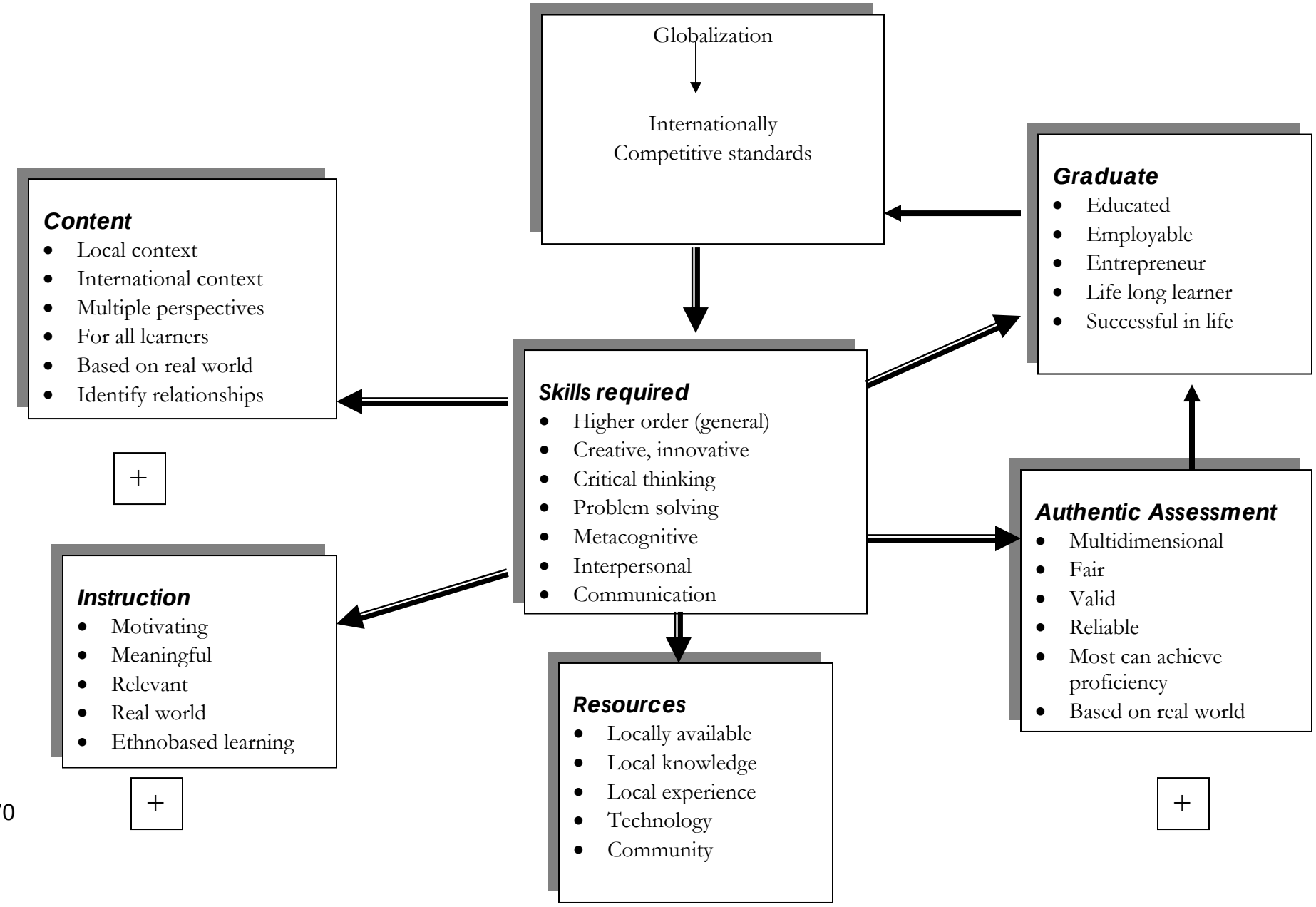
Items set for authentic assessment should:

1. be derived from real life situations experienced by the learner(s),
2. be based on experiences within and outside the school,
3. focus on issues that people care about or are important in their daily lives. The information must be credible and applicable to real–life situations such that they promote life-long learning. Any evaluations, conclusions and suggestions to be made must be for continuous improvement of life or must be towards sustainable living,
4. link local knowledge, culture, values and beliefs with other cultures including Western culture, technology and the contemporary world,
5. show equity by enabling all learners to demonstrate their academic achievements,
6. not be biased towards or against any group of learners,
7. avoid stereotyping, unreality, fragmentation, isolation and language bias,
8. although focusing on a subject, reflect content integration e.g. across topics, subjects etc. as experienced in real life,
9. not test recall, but ability to make sense of information provided and apply it,
10. assess ability to apply knowledge to real world situations,
11. enable the learner to demonstrate analytical, evaluative, critical thinking and problem solving skills,
12. measure abilities such as conceptual understanding, procedural knowledge, problem solving, linking, discerning and visualizing relationships, manipulating data to come up with new knowledge or materials to create new objects and application of computational skills to relevant situations.

References

- Alaska Curriculum Framework Project. http://www.edu.state.ak.us/tls/frameworks/langarts/2strtput.htm
- De Silva, S. (1997) 'Human Resources Development for Competitiveness'. Paper presented at the ILO Workshop on Employers' Organisations. Turin, Italy 5 – 13 May 1997
- Doolittle, P.E. (1997) "Constructivism and Online Education", Virginia Stake University.
- Mowl, G. (1996) ' Innovative Assessment"
- Pandey, T. (1990) "Authentic Mathematics Assessment" http://www.exit109.com/~learn/mathases.htm retrieved 02/08/26.
- Ruby A. (1996) "The Holy Grail of Education Reform; Internationally Competitive Standards". http://www7.nationalacademies.org/taking-stock-workshop-papers-ruby.htm
- Suarez-Orozco M.M and Qin-Hilliard (2004) "Globalization". http://www.ucpress.edu/books/pages/10166.cho1.html retrieved 2/23/2006.

Appendix: International standards of education model



Locating the local village within the global village

Assessment possibilities and practical challenges

Vanessa Scherman, Elizabeth Archer and Sarah Howie, Centre for Evaluation and Assessment, Faculty of Education, University of Pretoria, South Africa

Abstract

The Centre for Evaluation and Assessment (CEA) situated at the Faculty of Education, University of Pretoria in South Africa has been working collaboratively with the Curriculum, Evaluation and Management (CEM) Centre at the University of Durham in the United Kingdom on an assessment project since 2003. The CEM centre has developed a suite of monitoring projects catering for learners from primary school, through to A-levels. The CEA has been researching the feasibility of adapting and implementing two projects, one for the primary school and one for the secondary school, for the South African context. The instruments that were developed by the CEM centre are currently being used as baseline assessments in a number of countries, including Australia, New Zealand, Scotland and Germany. In contrast to these countries, South Africa is a developing country, with vast discrepancies in terms of schooling conditions and resources with the additional challenges of multilingualism in the classroom. These issues complicate the implementation of equitable assessment practices. The tension arises between adequately mapping the instruments in terms of context specific monitoring of achievement within South Africa, while maintaining the integrity of the instrument for the purpose of international comparisons. In this regard issues of validity, reliability, fairness and practicality are highlighted. These issues pertain to the quality of the instruments and the research question addressed is: **To what extent can an international monitoring system be adapted for the South African context and implemented effectively**. This paper addresses these issues as part of an ongoing research project, funded by the National Research Foundation (NRF).

Introduction

This paper aims to explore the possibility of using monitoring systems developed internationally for 'national monitoring' to a developing world context such as South Africa.

The guiding research question is: **To what extent can an international monitoring system be adapted for the South African context and implemented effectively**. Here certain issues come to the fore, namely to what extent are the assessments valid and reliable, the issue of equity is raised in addition to that of fairness and practicality.

These issues (validity, reliability and equity) are discussed against the backdrop of quality education. The challenge of any education system is to be able to provide quality education for participants in the system and it is not surprising that internationally there has been a reemphasis on quality education. Two of the recent United Nations conferences namely the Jomtien Declaration in 1990 and the Dakar Framework for Action in 2000 have recognised that the quality of education is imperative if goals and objectives of developing countries are to be met (UNESCO, 2005). However, there is little consensus on what quality education is, as the concept could be understood differently by different stakeholders (Fitz-Gibbon, 1996) and when asked to describe quality many would use the terms useful, good, efficient or measuring up (Botha, 2002).

In 2003, the Centre for Evaluation and Assessment (CEA) at the University of Pretoria in collaboration with the Curriculum, Evaluation, and Management Centre (CEM) at the University of Durham, in the United Kingdom, embarked on a research project to investigate the possibility of adapting existing monitoring systems established in the United Kingdom for the South African context. This project is funded by the National Research Foundation, a national funding body in South Africa, to investigate the possibility of adapting existing monitoring systems established in the United Kingdom for the South African context. The aim of adapting the monitoring systems is to provide information on the quality of education learners receive, quality here specifically referring to whether academic gains are made.

The CEM centre is a research centre in the United Kingdom and has developed a number of monitoring systems at various stages of the United Kingdom schooling system. Most well known are the Primary Indicators at Primary Schools (PIPS), Middle Years Information System (MidYIS), Year 11 Information System (YELLIS) and finally A-level Information System or ALIS (CEM, 2002a). Although there were several projects which could be investigated the CEA decided to focus on PIPS, which would be implemented at the beginning of primary school and MidYIS which would be implemented at the beginning of secondary school which were strategically the two grades most in need of baseline measures as they are the beginning of primary and secondary school (the South African version of PIPS is referred to as PIPSSA which is Primary Indicators at Primary Schools in South Africa, while the South African version of MidYIS is referred to as SASSIS which is South African Secondary School Information System).

The monitoring systems developed by CEM were feasible options as the CEA identified a potential need for monitoring systems for schools as there is no specific policy or programme in place to monitor or evaluate learners at the beginning of primary and secondary school and the CEA recognised that it was precisely at these levels collecting baseline information would be invaluable in order to track the progress of learners. There are currently a couple of policy initiatives that are relevant to the discussion on quality of

education and monitoring. These are the policy on Integrated Quality Management System (IQMS) (including the Whole School Evaluation programme) and Systemic Evaluation. As part of the Whole School Evaluation component of the Integrated Quality Management Systems schools should be evaluating themselves on a yearly basis and many schools felt ill-prepared to undertake the official self-evaluation programme under the IQMS. It was believed that schools would welcome a system of monitoring that would permit them to evaluate their learners and then track them as well as collecting other information (e.g. attitudinal). This would provide them with an insight into their learners, their behaviour and performance and therefore not only enhance the school's ability to intervene where weaknesses were identified but this would also provide them with important information ahead of the IQMS evaluation.

The paper begins with an outline of national monitoring systems developed in other countries (Section 2), followed by the rationale for why the monitoring systems developed by CEM were selected (Section 3). The concept of equitable assessments practices is then elaborated on specifically highlighting issues of validity, reliability and fairness. The paper concludes with recommendations on the way forward.

Monitoring in education

School success has often been thought of in terms of achievement and tools used to monitor progress of learners in order to ensure achievement (Safer & Fleischman, 2005). However, school success is not just achievement and the concept of monitoring needs to be defined whilst presently there is little agreement in literature on the definition of monitoring (Sammons, 1999). Even though there is little agreement of what the concept means, monitoring is constantly mentioned in school effectiveness research and is often linked to the achievement of learners (Scheerens, Glas & Thomas, 2003:14):

> …frequent monitoring and evaluation of learners' progress stands out as a factor that is consistently mentioned in research reviews as a correlate of educational achievement.

Scheerens *et al* (2003) are of the opinion that monitoring can be defined as a systematic gathering of information in order to make judgments about the effectiveness of schooling. Furthermore, monitoring stresses ongoing gathering of information as a basis for making decisions. Raffan and Ruthen (2003) further elaborate on the gathering of information by linking the activity to learning and keeping an eye, if you will, on learning in terms of difficulties experienced and progress made or in other words focusing specifically on the learner and classroom level providing a mechanism of formally regulating the desired level of quality (Scheerens *et al*, 2003) by means of informed planning, teaching and assessment. Monitoring assesses achievement trends over time (Lockheed, 1996) and in the words of Hager and Slocum (2005: 58) "a system for ongoing progress monitoring is critical to ensure the student is continually moving toward mastery." For the purpose of this paper monitoring is seen as gathering relevant information on learner performance at various stages in order to ascertain whether academic gains have been made in order to identify strategies were necessary (Scherman, 2006).

There are a number of monitoring systems internationally that illustrate the characteristics required of good monitoring programmes, namely that the monitoring system includes a manageable unit of education, has an explicit rationale underpinning the system as well as a primary aim, is negotiated among stakeholders and has a positive affect on behavioural aspects, as well as should not interfere with the system that is being monitored (Fitz-Gibbon, 1992). These monitoring systems include the ZEBO-project developed in the Netherlands. The ZEBO-project that consists of three elements namely a pupil monitoring system, (ZEBO-PM), an assessment of educational content covered (ZEBO-CC) as well as measures of school process indicators (ZEBO-PI) (Hendriks, Doolaard & Bosker, 2001; Hendriks, Doolaard & Bosker, 2002). Also, in Australia the Victorian Certificate of Education data project can be identified which aims to assist schools to monitor effectiveness of teaching and learning in 53 subjects over a period by providing schools with performance data (Rowe, Turner & Lane, 2002). While in the United States of America the ABC+ (Attitudinal/ Behavioural/ Cognitive Indicators plus Context) model can be identified which aims to provide process data to schools and districts at the classroom-, grade-, and school level in order to develop school improvement plans that are driven by best practices in school effectiveness and staff development research (Teddlie, Koshan & Taylor, 2002). A similar project to the ABC+ model is the Assessment Tools for Teaching and Learning or asTTle that was developed in New Zealand which also focuses on school improvement. asTTle aims to provide educators with a resource which will assist in the creation of tests in reading, writing and mathematics, includes an input function for performance as well as national norms and comparisons to cohort groups but perhaps more importantly provides diagnostic information for individual learners and the class as a whole. The diagnostic information can then be used for future teaching based on the strengths and the weaknesses of learners (Ward, Hattie & Brown, 2003). Finally, the Tennessee Value-Added Assessment System (TVAAS), which was developed in the United States, can be identified. This system is called a value-added system where value added refers to a model in which academic gains made by learners are investigated. The primary purpose of TVAAS was to provide information for summative evaluations pertaining to how effective a school or educator has been in leading learners to achieve academic gains over a period of time (Sanders & Horn, 1998), reflecting growth regardless of initial levels of performance (Sanders, Wright, Ross & Wang, 2000).

A monitoring system for South Africa

A monitoring system in the South African context has to serve the same purposes as the examples briefly discussed thus far. The lessons for South Africa that can be taken from these examples are clear in that it is pertinent to consider that although the classroom and the school-level are the primary focus, other areas of the system such as the district and provincial level cannot be ignored. Thus, one has to consider the inclusion of the parents or community in addition to higher levels of the education system such as district, provincial or national level. Furthermore, the rationale has to be clear in that is the goal to develop tools for self-evaluation to monitor effectiveness or is the goal to make use of already developed tools in order to develop self-improvement plans. Finally, the level of participation of the school has to be identified, does the school collect the information themselves, send the information for capturing and transformation and then analyse the data or does the school liaise with research consultants who collect the data, analyse the data and provide detailed feedback reports.

For South Africa, and in light of policy initiatives, it would be important to include other levels of the system as well so as to ensure that vital elements within the system are included. For example, without inclusion of the district office schools may not be able to obtain the support they need to carry out improvement plans. Furthermore, in light of the uncertainty as to what is expected in terms of self-evaluation as well as the timelines associated with self-evaluation processes, it may be beneficial to make use of instruments which are already developed but can be adapted to the South African context. As this may take the least time in terms of development but could potentially yield effective results. Finally, with the demands placed on schools it is not likely that they will have the time to collect and analyse the information themselves but rather make use of researchers who will be able to collect the necessary data as well as supply the information that is needed tailored, to the school's needs.

The CEM centre has also developed a number of monitoring systems using value-added systems at various stages of the United Kingdom schooling system as mentioned earlier. Not only has the CEM centre developed monitoring systems at ever level of the schooling system, the Centre also enjoys substantial support from the educational community and the schools in the United Kingdom particularly schools pay for the services offered by CEM. Furthermore, the monitoring systems developed by CEM are what one would call a ground-up approach as schools have chosen to participate in the projects. This approach is in contrast to the top-down systems that are imposed on schools by the Education system. Moreover, the development of the monitoring systems was determined by the need to measure outcomes along with covariates so that fair comparisons can be made as well as process variables from which hypotheses could be generated. This approach invariably is appealing especially considering South Africa's apartheid past and now where there is a need to make fair comparisons and equitable assessment practices.

The South African Qualifications Authority (SAQA) specifically aims to establish equitable assessment while remaining cognisant of the history of South Africa and the reconstruction and development goals of the new democratic government of this country as well as the need to align the South African education and training system to emerging international trends of best practice in the provision of quality education and training and lifelong learning (SAQA, 2001). Thus the South African version of adaptation of the PIPS and MidYIS instruments from the CEM centre would have to be aligned with equitable assessment practices through accommodations for the unique South African context. Equitable assessment practices are discussed in the following section.

Equitable Assessment Practices

According to Borg (2001) equitable assessment allows "for learners (i) who learn in different ways, such as we see in multiple intelligence theory, (ii) who have different backgrounds which act as unique learning frameworks, (iii) who may be at different developmental stages and (iv) who develop a different understanding of the instructional process, such as a learning difficulty or lateral thinking." This clearly illustrates the wide range of concepts incorporated in equitable assessment.

The National Center for Research on Evaluation, Standards and Student Testing (CRESST) (1999:1) in the USA defines equity in assessment as follows:

> Equity is the concern for fairness, i.e., that assessments are free from bias or favoritism. An assessment that is fair enables all children to show what they can do. At minimum, all assessments should be reviewed for (a) stereotypes, (b) situations that may favor one culture over another, (c) excessive language demands that prevent some learners from showing their knowledge, and (d) the assessment's potential to include learners with disabilities or limited English proficiency.

The suite of instruments developed for the CEM centre was developed and specifically designed for the English context. As the South African context differs widely from that in the United Kingdom the unique learning context of South Africa may influence how children perform on these instruments. Certain accommodations and adaptations of the PIPS and MidYIS instruments were thus needed in order to develop instruments, which would provide equitable assessment information. Furthermore, the idea of equitable assessment is encompassed in the SAQA principles of good assessment namely fairness, validity, reliability and practicability (SAQA, 2001: 16). Each of these principles is discussed separately in the following sections.

Fairness

SAQA explains fairness as taking account of and addressing of issues pertaining to the inequality of opportunities, resources and appropriate teaching and learning approaches in terms of acquisition of knowledge understanding and skills. Here issues of bias in respect of ethnicity, gender, age, disability, social class and race in the assessment approaches, instruments and materials are important. In addition, what is being assessed has to be clear (SAQA, 2001).

The idea of fairness in equitable assessment obviously stretches well beyond only cultural fairness. Fairness in assessment is often accomplished through accommodations to an existing assessment where adjustments are made in terms of settings and procedures or controlling of intervening factors such as culture, which complicate the assessment of a specific construct. Effective accommodations and adaptations boost the performance of learners influenced by these intervening factors, but not that of learners unaffected by these factors (Elliott, McKevitt and Kettler, 2002; Bowen & Ferrell, 2003; Thompson & Ouenemoen, 2003; Fuchs, Fuchs, Eaton, Hamlett & Karns 2000). A multitude of accommodation possibilities have been highlighted by authors (Bowen & Ferrell, 2003; Hofstetter, 2003; Polloway, Epstein & Bursuck, 2003; Eliott et al., 2002; Taylor et al., 2002, Ysseldyke et al., 2001), these accommodations to establish fairness encompass changes in scheduling, setting, equipment or technology, presentation and response.

Issues of validity and reliability are intrinsically related to appropriate accommodations and adaptations of assessment. Eliott et al. (2002: 155) sees accommodations as providing access to the instrument and assessing a child without exposure to social practices would thus translate into inequitable assessment practice as the child will be limited in the use of reading strategies such as reading for meaning and utilising context clues. Any adaptation of the CEM centre instruments to the South African context would thus need to consider of exposure to specific contexts and cultural practices and even types of representation to ensure that these do not act as intervening variables and thus undermine the validity of the instrument in the South African context by confounding the underlying constructs being examined.

Issues of validity

The nature of this paper precludes a comprehensive discussion on this aspect, but a fuller and more detailed discussion may be found in Scherman (2006). What follows here is a brief summary.

The central validity issue in adapting an assessment for the South African context is determining which adaptations and accommodations would preserve the meaningfulness of the scores (Fuchs, et al. 2000: 66). When accommodations produce scores for children in South Africa, which measure the same attributes as the original assessment measures for children in the country for which it has been developed, the instrument can be said to be valid for the South African context. It is thus the removal of the irrelevant construct variance created by the difference in culture, context, language, social practices, etc. which results in validity.

According to Cohen, Manion and Morrison (2003: 105), validity is basically the view that "…a particular instrument in fact measures what it purports to measure…" Validity addresses the question: to what extent is the interpretation of results appropriate as well as meaningful (Gronlund, 1998), and is a unitary concept that is based on various forms of evidence, with construct-related validity being the central concept, and ultimately is concerned with the consequences of using the assessment or questionnaire (Gronlund, 1998; Linn & Gronlund, 2000). Under the unitary concept face validity and content-related validity can be identified.

Face validity or the superficial appearance of what the test measures from the perspective of the participant is subsumed under content-related validity (Urbina, 2004). While content-related validity is generally understood as the extent to which how well the questions in the assessment matches the field within which the assessment can be located (Coolican, 1999). Thus, the sampling of items from the broader domain and items included is important (Gronlund, 1998) in terms of relevance as well as representativeness (Urbina, 2004). Content-related validity, which includes face validity and curriculum validity, where curriculum validity refers to the extent to which the abilities or competency assessed matches the curriculum (Thorndike, 1997), is established by means of drawing up tables of specifications or by consulting content specialists (Suen, 1990).

Reliability

Generally, reliability refers to the consistency of scores, which are obtained by the same individuals when they are requested to complete the assessment on different occasions (Anastasi & Urbina, 1997). Furthermore, reliability is important, as unless results are stable one cannot expect the results to be valid. Reliability not only gives an indication of how much confidence can be placed in a particular score obtained but also how constant the scores will be which are obtained in different administrations (Owen & Taljaard, 1996).

The consistency gives an indication of the ability of items to measure the same variable or construct where inconsistent items do not measure the same construct. Internal consistency is used in this study and is a pre-requisite for construct validity, where one would expect a

high item-total correlation since items measuring the same construct contributes to the total score of a test (Kline, 1993).

Issues of practicability

According to SAQA practicability refers to taking available financial resources, facilities, equipment and time into account. This speaks directly towards issues of sustainability and is very closely related to the specific context where the instrument is being employed. Any accommodations and adaptations have to be closely related to the instructional approach and material utilized in the classroom. If accommodations are unfamiliar to learners such as the use of technology or specific presentation mediums, the accommodations and adaptations in themselves can decrease the performance achievement of the learners on the assessment (Ysseldyke et al., 2001; Elliot et al., 2002; Wasburn- Moses, 2003).

The CEA through its knowledge and experience of assessment and evaluation in the South African context sought to adapt the CEM centre instruments to be as valid as possible for the South African population. The specific experience in the adaptation of the PIPS and MidYIS instrument to the South African context is discussed below.

Research design

A mixed methods approach was followed in this research, namely the integration of both quantitative and qualitative methods. This provided the researchers with additional opportunities for answering the research questions adequately (Teddlie & Tashakkori, 2003). A detailed discussion of the design and methods used may be found in Scherman, 2006.

Sample

Several schools were purposefully selected to participate in this project for maximum variation in their characteristics and background. As the aim of the research is to develop a monitoring system, which would be appropriate for South African schools regardless of the variation in schools, it is was imperative to include schools from various backgrounds. Due to financial constraints a limited number of schools could be accommodated. The sampling for the PIPSSA and SASSIS projects is discussed below.

PIPSSA

The PIPSSA project sampled seven schools of which four were former White schools[1], two of these were English medium schools and two Afrikaans medium, two of the schools were from the former African while one school from the former Indian participated. Schools generally requested that most of their Grade 1 classes be included, thus between one to four classes were assessed per school depending on the needs of the specific institution. In total, 426 learners participated in the 2005 PIPSSA study, of which the average age was 7 years (minimum age 6 years and maximum age 8 years) and 54% were male.

SASSIS

[1] For the study, due to financial constraints schools were selected in order to provide maximum variation in order to see if the instruments were suitable. White schools are the former Model C schools while African schools are the former Department of Education and Training. Indians schools are the former House of Delegates while Coloured Schools are Former House of Representatives.

For SASSIS three former White schools of which two were English medium and one school dual medium were included as well as three African schools, two Indian schools and finally two Coloured schools. Two classes from every school were randomly selected[2] by means of WinW3S. Thus, all learners had an equal and independent chance of being selected (Gay & Airasian, 2003). In total 794 learners participated, of which the average age was 14 (minimum age 12 and maximum age 19) and 51% were female.

Instruments implemented in this research

PIPSSA

The PIPSSA assessment is computer-based and was loaded onto laptop computers and, with the help of trained fieldworkers, administered to learners at participating schools. The PIPSSA instrument consists of 17 subtests, which are combined into three different scales: early phonics, early reading and early mathematics. The scales are generated as follows:

SASSIS

The assessment instrument for SASSIS is paper-based and consists of seven subsections which were collapsed into four different scales namely the vocabulary scale, the mathematics scale, the skills scale, and the non-verbal scale each of which were designed to measure certain skills and abilities (the scales and the subtests are discussed below). The seven subtests were timed and consist of multiple-choice items with the exception of the mathematics subsection, which included both constructed response items as well as multiple-choice items. The scales are:

1. The Vocabulary scale is derived from the subtest with the same name in the assessment and measure abilities in vocabulary as well as fluency and speed.
2. The Mathematics scale is derived from the subtest with the same name in the assessment and measure abilities in mathematics as well as fluency and speed.
3. The Skills scale comprises two subtests namely the Proof Reading subtest and the Perceptual Speed and Accuracy subtest. The Proof Reading and Perceptual Speed and Accuracy subtests are designed to measure fluency and speed in finding patterns and spotting mistakes and as such rely heavily on the learner's scanning and skimming skills.
4. The Non-Verbal scale comprises three sections namely Cross Sections, Block Counting and Pictures. These tests attempt to measure 2-D and 3-D visualisation, spatial aptitude and pattern recognition. The Non-verbal score is a useful indicator of ability for learners for whom English is a second language, as there is no reliance on language (CEM, 2002b).

The assessment is a combination of a speed assessment and power assessment where a speed assessment measures the speed with which participants perform tasks and the difficulty of tasks are manipulated through timing. While, a power assessment on the other hand has no time limit and difficulty is manipulated by increasing or decreasing the level of complexity of items. As the assessment is a combination of a speed assessment and a power assessment, the time limits typically allow the majority of participants to attempt most or all of the items (Urbina, 2004).

[2] WinW3S was used for this and it is a within-in school sampling package developed by the Data Processing Centre of the International Association for the Evaluation of Educational Achievement (IEA). Special permission was obtained to use the program as the program is normally only used in IEA studies.

Data Collection

PIPSSA

For the PIPSSA component, laptops were used for the data collection. A team of between 4-6 people went out into the schools after being trained on how to operate the software and how to conduct an assessment. Each fieldworker assessed one learner at a time. The fieldworkers were given a venue (whether it is an unused classroom or the school hall) to then assess the learners. The use of the computer-based assessment meant that standardised procedures could easily be followed as the assessment was guided by the program itself. Assessments took place in English, Afrikaans and Sepedi, depending on the language of teaching at each school. The learners were assessed and the data was captured immediately onto the computer in the form of DAT Text files. Once the data was collected, the data was downloaded from the laptops.

SASSIS

Each school was visited on a separate day and fieldworkers administered the instruments. Each classroom had a fieldworker overseeing the standardised administration procedure. The fieldworker read a script explaining the assessment and questionnaire as well as the time limits for each subsection. This ensured that the administration procedures were standardised across the schools and that each learner receives exactly the same information. The assessment as well as the questionnaire took approximately two and a half hours to complete. The English script was translated into Sepedi and Afrikaans (the two additional language of instruction for the sampled schools) in order to ensure that each learner would understand what was expected. Two groups of translators were used for the translation of the administration script. The first group translated the English script into Sepedi and Afrikaans while the second group of translators checked the Sepedi and Afrikaans translations against the English version. Any changes or corrections were made and the scripts finalised. Thus, administration of the assessments took place in English, Sepedi and Afrikaans depending on the school that was visited.

In order to capture the administration process the fieldworkers completed an administration questionnaire detailing the administration process, which includes problems experienced, comments made by learners and general impressions as well as time taken for the majority of learners to complete the subsection.

Data analysis

Document analysis

The document analysis for both the PIPSSA and SASSIS instruments included examination of the curriculum policy documents, specifically the Language Learning Area and the Mathematics Learning Area curriculum documents. The documents were imported into Atlas *ti,* and analysed over a two-week period, the results were used in conjunction with the evaluation reports by expert evaluators.

Analysis of the validity and reliability of instruments

In order to investigate the different aspects of validity (in this case face and content-related validity) specialists in the field of psychology and education were approached. Two research

psychologists as well as an educational psychologist evaluated the assessment instrument for content-related validity. The psychologists were asked to complete an evaluation form. A meeting was scheduled to discuss the results of the evaluation.

Furthermore, specialists in the field of education, specifically mathematics and language, were also approached and the assessment was evaluated from a curriculum perspective. The specialists were asked to complete an evaluation form in addition to drawing up a table of specification. Once the evaluation task was completed, a meeting was scheduled with each specialist to discuss the results of the evaluation.

Internal consistency reliability[3] was used in the analysis, which is a pre-requisite for construct validity, where one would expect a high item-total correlation (above .70) since items measuring the same construct contributes to the total score of a test (Kline, 1993). Over and above indicating the stability of measures over time, this would also strengthen the inferences that could be made by the researchers on the content-related validity of the assessment (Suen, 1990).

Results

Document analysis

Initial indications are that it would appear from the policy documents that there is a reasonable overlap between the assessments for both primary and secondary school and the intended policy documents. The overlap for the primary school components was better than the overlap of the secondary school component. The result is perhaps not surprising as the primary school assessment was developed with a curriculum in mind while the secondary school assessment was developed as an "abilities assessment".

In the primary school assessment the Early Phonetics scale relates to the objectives of Listening Reading and Viewing as well as Language as denoted in the Revised National Curriculum for Grade-R and Grade 1 (National Department of Education, 2002a). The Early Reading scale addresses the outcomes as set out in the Revised National Curriculum for Grade-R and Grade 1 such as understanding the purpose of print, distinguishing letters, awareness of directionality and the ability to identify words. The Early Mathematics scale addresses the outcomes of the Revised National Curriculum for Grade-R and Grade 1 (National Department of Education, 2002b). The following skills are assessed: numbers, operations, relationships, space, quantity, counting, simple calculation, working with money, fractions, simple division and shape as well as measurement.

For the secondary school assessment it was found that the type of skills assessed was present in the language and mathematics curriculum. Of the six outcomes in the language curriculum, three learning outcomes are highlighted namely *Listening* as learner have to listen

[3] Generally, reliability refers to the consistency of scores, which are obtained by the same individuals when they are requested to complete the assessment on different occasions (Anastasi & Urbina, 1997). Consistency gives an indication of the ability of items to measure the same variable or construct where inconsistent items do not measure the same construct.

to the instructions, *Reading and Viewing* as well as *Language Structure and Use*. The learning outcomes mentioned correspond to the Instructions of the assessment as well as Proof Reading and Vocabulary where Proof reading corresponds well with *Reading and Viewing* as well as *Language Structure and Use* and Vocabulary with *Language Structure and Use*. There is a greater overlap with the mathematics curriculum and the secondary school assessment as four of the five outcomes are represented, namely Numbers, Operations and Relationships, Patterns, Functions and Algebra, Space and Shape, Measurement. The four learning outcomes correspond well with Perceptual Speed and Accuracy (Patterns, Functions and Algebra), Mathematics (Numbers, Operations and Relationships, Patterns, Functions and Algebra as well as Measurement), Block Counting (Space and Shape), Pictures (Patterns, Functions and Algebra) and Cross Sections (Space and Shape). However, before any final decisions can be made the results from the expert evaluation has to be considered.

Expert appraisals

The specialists in Education from mathematics and languages were approached to evaluate the assessments for both primary and secondary school. The same brief was given to the specialists namely to evaluate the subtests and items in terms of content validity as well as curriculum validity. The specialists were asked to develop assessment frameworks to match items to learning outcomes. The results are given separately for the two components below:

PIPSSA

The external reviewers found that there was comprehensive overlap between the curriculum and the instrument. The mode of presentation of the items, in terms of using laptops however leads to some difficulty in the PIPSSA project.

1. Financial demands. Most of the schools involved in this project do not have computer laboratories; as such laptop computers need to be rented for the fieldwork. The cost of renting laptop computers for the fieldwork represents a major part of the expenditure in this project.
2. Security. Travelling with valuable equipment such as laptops present a serious security risk in South Africa. This may negatively impact the safety of the fieldworkers.
3. Administrative burden. The process of booking, renting collecting and returning the laptop computers as well as having to repeatedly upload the necessary software for fieldwork represents a large administrative burden to the CEA team. This translates into many person-hours of labour, which may have been used more productively.
4. Administration time. The administration time of twenty minutes projected per child (Tymms & Wylde, 2003) is greatly increased in the PIPSA project as laptops must be set up and fieldworkers are often not as computer literate as the educators in the UK.
5. Sustainability. In order to achieve true sustainability for this project it would be necessary to empower educators to administer this test and relay the data to the CEA. In order to achieve this, it would be essential to ensure that the necessary infrastructure is in place. Currently there are vast discrepancies in the availability of computer facilities for schools in South Africa. The South African Department of Educations Draft White Paper on e-education of August 2003 indicates that in 2002 only 26.5% of schools had access to computers for teaching and learning. The availability of computers for educational purposes varied from only 4.5% in the Eastern Cape to 56.8% in the Western Cape. Although some provinces have launched ambitious programmes to equip all schools in the province with computers such as

the Gauteng Online project, the targets set have not yet been realised (Gauteng Department of Education, 2005). It may be more prudent to follow the same course of action as the American Dietetics Association which delayed the switch to computer based assessment due to inaccessibility of test centres with computer facilities until such a time that they had established an appropriate infrastructure (Ruiz, Fitz, Lewis & Reidy, 1995).

Some of the items in the assessment were however deemed as being too Euro-centric by external evaluators. The expert evaluators of the primary school assessment made suggestions to make the assessment more appropriate for the South African context. Examples of suggestions for the primary school component included:

1. The expert evaluation reports indicated that the computer-based PIPS was likely to disadvantage learners who have not been exposed to cartoons, animations and three-dimensional overlays. These elements are less pronounced in the paper-based version of the PIPS assessment. Learners may be distracted or misled by the graphic representations and the assessment would thus not truly be assessing what it purports to.
2. The reviewers indicated that some of the graphic representations were very Euro-centric and may have to be replaced with more South African representations. For example replacing the beach balls with soccer balls.
3. The reviewers found some of the phonetic items inappropriate for all language groups as the pronunciation amongst various language groups can differ widely and diagraphs and diphthongs are often found in different placements in African languages than in English.
4. The reviewers indicated that the vocabulary section would have to be revisited. The specifically indicated items such as gnome, toadstool, castle and cherries as possibly being inappropriate for the South African context.

SASSIS

For the secondary school component the task of matching the curriculum and the assessment was easier to accomplish for mathematics then for language. Although it would be possible to construct a similar table for the Vocabulary subtest which forms part of the Language Learning Area it is more complicated for the Proof Reading subtest as the learners are provided with a passage that they have to correct and not singular items which can be neatly characterised as easy, moderate or difficult. It is for this reason that a similar table for the Language Learning Area is not provided.

The analysis of overlap between the Mathematics Learning Area and the SASSIS assessment proved to be very fruitful. The mathematics specialist indicated that skills needed for four out of the five learning outcomes where represented in the assessment namely Learning Outcome 1: Numbers, operations and relationships, Learning Outcome 2: Patterns, functions and Algebra, Learning Outcome 3: Space and shape and finally Learning Outcome 4: Measurement. The specialist however raised a concern that certain items were excessively easy, that Learning Outcome 1 and 2 were over represented in the mathematics subtest of the assessment and that the time limits needed to be revised. Furthermore, the mathematics specialist indicated that certain items were not present in the mathematics curriculum but that the items would be accessible to an average Grade 8 learner as a result of general knowledge, experience and problem solving strategies.

Upon analysing the content of the assessment, the language specialists indicated that the Instructions, Vocabulary subtest, and Proof Reading subtest were of relevance for the Language Learning Area in that the skills assessed are taught in the curriculum specifically Learning Outcome 1: Listening, Learning Outcome 3: Reading and viewing and Learning Outcome 6: Language structure and use. Furthermore, one of the specialists indicated that the items were not bias in terms of gender or race and that the language used is age appropriate. However, the other specialist indicated that although the basic skills are present in the curriculum that certain items would prove difficult for second language learners and that these items should be evaluated.

In order to make the assessments for both primary and secondary school relevant for the South African context suggestions included in the expert evaluations were effected before data collection. Examples of suggestions for the secondary school component included:

1. The expert evaluation reports indicated that the instructions could be ambiguous and difficult to follow. Thus, the instructions were rewritten so that learners would understand what was expected of them but that the rewritten version would still be comparable to the original.
2. The reviewers indicated that should a learner be unsure of what to do that they would have to page to the beginning of the subtest in order to reread the instructions. This wastes time. Thus, the instructions were included at the top of the page throughout the assessment so that learners if uncertain could reread the instructions without wasting time.
3. The reviewers were not happy with the time limits allocated to the subtests however the majority of the learners were able to complete 90% of items per subtest with the exception of Mathematics and Proof Reading. Therefore, the time allocated for each subtest was increased so that the majority of the learners would be able to or almost be able to complete the subtest.
4. The reviewers indicated that certain words in the vocabulary section were ambiguous and that the way in which the words were presented was not in line with how vocabulary was taught in South Africa. As a result, the vocabulary subtest was revised not only were ambiguous words replaced but also the core word for which a synonym had to found was placed within the context of a sentence. It is suspected that the as a result the items may be easier but more accessible to second language learners.

In addition to specialists in Education, specialists in the field of Psychology were also asked to evaluate the assessment from a psychological assessment point of view. The brief was to review the instruments for content-related validity. An Educational psychologist as well as Research psychologists formally reviewed the instrument. The outcome of the reviews indicated that the subtest do correspond with the domain of items found in ability assessments.

To conclude, the decision of the education specialists indicated that the assessment was relevant for the South African curriculum although certain changes would need to be effected. Additionally, the specialists in Psychology indicated that the items included in the assessment adequately sampled the domain of abilities assessments.

Reliability analysis

PIPSSA

Reliability analysis was undertaken for 16 of the subtests (excluding the handwriting subtest) as well as the three scales of the assessment (Table 1 and Table 2). As can be seen the reliability coefficients are quite high all above .79 with the reliabilities reported for the Letters (0.97) and Stories subtest (.96) as well as the Early mathematics scale (0.96) and Early reading scale (0.95) The reliability coefficient for Shapes was by far the lowest with .79. Vocabulary and Reading items, which were deemed inappropriate for the Afrikaans and Sepedi learners, were not included in the analysis.

Table 1 Reliability coefficients for the fifteen PIPSSA subtests

Subtest	Cronbach Alpha
Rhyming Words	.85
Repeating words	.83
Vocabulary	.93
Ideas about reading	.85
Letters	.97
Mix up words	.88
Quiz words	.92
Stories	.96
Sentences	.91
Sizes	.87
Counting	.82
Sums A	.85
Numbers	.91
Shapes	.79
Maths	.86
Sums B	.95

Table 2 Reliability coefficients for the three PIPSSA scales

Scale	Cronbach Alpha
Early Reading	.95
Early Phonics	.86
Early Mathematics	.96

Table 3 Reliability coefficients for the three PIPS scales for the UK

Scale	Cronbach Alpha
Early Reading	.97
Early Mathematics	.90
Total	.98

(CEM, 2002c)

As can be seen from the Table 3 the reliability coefficients for South Africa compare well with the reliability coefficients for the United Kingdom. The results are encouraging as although the assessment needs to be developed further in terms of face validity the items themselves are sound for our context.

SASSIS

Reliability analysis was undertaken for the seven subtests as well as the four scales of the assessment (Table 4 and Table 5). As can be seen the reliability coefficients are quite high all above .71 with the reliabilities reported for the Perceptual Speed and Accuracy and Proof Reading subtest (.94) as well as the Skills scale which comprises the two subtests (0.95). The reliability coefficient for Cross Sections was by far the lowest with .71. Upon inspection and exploring the item statistics it was found that two items were problematic and was removed from the analysis. The two items in question were similar in nature and as a result caused some confusion.

Table 4 Reliability coefficients for the seven SASSIS subtests

Subtest	Cronbach Alpha
Vocabulary	.91
Mathematics	.91
Proof Reading	.94
Perceptual Speed and Accuracy	.94
Cross Sections	.71*
Block Counting	.78
Pictures	.82

* After Two items deleted

Table 5 Reliability coefficients for the four SASSIS scales

Scale	Cronbach Alpha
Vocabulary	.91
Mathematics	.91
Non-verbal	.88
Skills	.95

Table 6 Reliability coefficients for the four MidYIS scales for the UK

Scale	Cronbach Alpha
Vocabulary	0.90
Mathematics	0.93
Non-verbal	0.89
Skills	0.84

(Source CEM, 2002c)

A similar trend can be seen for the MidYIS/SASSIS component of the project as with the PIPS/PIPSSA component in that the reliability coefficients for both South Africa and the United Kingdom are comparable. Thus one could tentatively conclude that the items included seem to be consistent across different contexts.

Discussion

Monitoring systems are important mechanisms that schools can use to gauge their effectiveness. If quality education is to be investigated then some form of monitoring is needed. The type of monitoring system used depends on the aim, purpose, or the rationale of the system. In section 2 of the paper several national monitoring systems from other countries was briefly described. What is clear from the brief description is that the monitoring systems had a clear aim or purpose. At the heart of the systems described was to

provide accurate information upon which decisions for teaching and learning could be based. For this research, the aim is not all that different in that the aim is to develop a system, which schools and educators could use to monitor learner performance as well as be used as a tool for internal evaluations, for improvement purposes.

However, monitoring systems implemented by schools and used to assist in self-evaluation processes in the context of South Africa are not available. The schools within South Africa vary greatly and schools within rural areas and townships are still disadvantaged in terms of resources and facilities. However, current assessments do not reveal the complexities within which disadvantaged schools work and in order to evaluate the true performance of a school more appropriate monitoring and measurement systems are necessary. Moreover, with the increasing demand of the provincial and national education departments to ensure that schools become accountable for their learners' performance, the need for a system, which monitors learner performance, has become imperative. Schools will have to develop the capacity to monitor their own effectiveness in order to be accountable for their learners' performance. By means of using systems such as PIPSSA and SASSIS with adaptations for the South African context school processes as well as outputs can be monitored.

Furthermore, equitability is a matter of degree and is achieved through a combination of fairness, validity, reliability and practicability. Explorations into the feasibility of using existing monitoring systems are promising especially when evaluating in terms of validity, reliability, and fairness, which are discussed separately.

Validity per se is inferred from evidence as well as ultimately depends on many different types of evidence from which inferences are drawn and expressed by degree such as high, moderate and low and is specific to a particular use. In terms of the evidence considered the validity of these assessments could be said to be moderate as all though there is considerable overlap between the content domain and the assessment for both PIPSSA and SASSIS there is also room for improvement, highlighted perhaps by the Euro-centric nature of diagrams included in the PIPSSA assessment as well as by the overlap of skills taught in the curriculum and the skills tested in the SASSIS assessment. However, considering that the assessment was developed in another country with similar and well as different objectives in mind the result is heartening especially as is that the use of the value-added approaches, as is used in these systems, contributed to establishing fairness and validity Furthermore, issues pertaining to validity of research are an important aspect more so now that there are a variety of methodological choices available. According to Newman *et al* (2003):

> …researchers strengthen validity …when they can show the consistency among research purposes, the questions and the methods they use. Strong consistency grounds the credibility of research findings and helps to ensure that audiences have confidence in the findings and implications of research studies.

A key issue, and a discussion, which has been taking place in South Africa for a while now is that of fairness particularly in terms of cultural fairness. As previous mentioned the assessments particularly the PIPSSA assessment, may not be as culturally fair in South Africa as it would in other another country perhaps. However, the assessment is accessible to

learners and a key challenge and part of the recommendations from the project team is to adapt the assessment so that it is not bias in terms of culture.

Furthermore, the reliability of these assessments, which ascertains the consistency of the results and gives an indication of how stable the results are were comparable to the results in analyses undertaken in the United Kingdom. Reliability coefficients should be high, above .7 for assessments. From the analysis undertaken it appears as if both the PIPSSA assessment and the SASSIS assessment are reliable as the coefficients are above 0.7. The result is perhaps not surprising as these are well-established assessments. However, the result contributes to the assertion that the assessments are valid for the South African context, thus positioning South Africa within the international arena.

The issue of practicality though has not been directly addressed thus far however; these have implications for the way forward. Practicality, in light of the current discussion, is thought of as the financial resources, facilities, equipment and time. Presently, the project is funded by the NRF so there are no financial resources from the school per se are required. However, in the future this may be an important consideration for school participation. Equipment and time are considerations however, especially in terms of the PIPSSA assessment. The assessment is a computer-based assessment and the participating schools do not always have the computers available and so far the CEA has been making use of laptop computers. The use of laptop computers also makes it possible for the project team to load the assessment and download the data at the end of every testing session. However, this is a time consuming and often a laborious task. Furthermore, the PIPSSA assessment takes anywhere between 25 minutes and 45 minutes per child to complete. This means that certain children are taken out of the classroom for extended periods of time, which is not always ideal.

Way Forward

This research is in progress and the data are not yet fully explored. Nonetheless, there are number of plans underway for the near future PIPSSA and SASSIS and these include:

1. **Further development of the monitoring system to include contextual indicators.** Different inputs, process and outputs should be included if the monitoring system is to be comprehensive in nature as well as tap different domains such as affective, cognitive, and behavioural. However, if the monitoring system is to be comprehensive then information from more than one level should be included. Thus, additional contextual information will have to be collected from the learner-, classroom- and school-level. The system in its present form only provides learner-level information. Thus, questionnaires will have to be developed and evaluated to ensure validity.
2. **Extended exploration of construct validity:** Problematic items have to be determined as well as underlying data structure to evaluate construct validity to ensure that the constructs or scales in the assessment are sound. Thus, factor analysis will have to be undertaken in order to ensure that the items in each subtest are testing the same construct. Furthermore, for the PIPSSA assessment a move back to the paper-based version is suggested as the graphical presentations in the paper based version makes use of less three-dimensional overlays that may distract from the construct validity of the instrument. Many of the graphic representations still maintain a cartoon or animation characteristic in the paper-based format all graphic representations will be changed in

terms style as changing only the most problematic items would result in inconsistencies in the style of representation throughout the assessment.

3. **Predictive validity has to be established for the South African context**: The assessment is used for prediction purposes in the context of the United Kingdom. If predicative validity is to be established for South Africa, the results from the assessment will have to be correlated with academic results, specifically language and mathematics, obtained from school-based assessments.
4. **Analysis procedures to be undertaken**: Analysis procedures used to provide information given to schools would have to be evaluated and appropriate analysis procedures for the initial validation phase as well as more developed phases will have to be identified. For example, standardised feedback cannot be given initially, as the assessment has not been standardised for the South African context and currently due to financial constraints and as a result small sample sizes the standardisation will not take place in the initial stages of the project. However, the aim is to standardise the assessment for the South African context and to develop national norms.
5. **The feedback reports to schools**: The feedback provided would need to be simplified and narratives added so that the results are presented in a comprehensive manner. Individual school reports are more appropriate in a South context that is presented to the schools during information sessions as well as follow-up telephone calls. The report should include background information on the assessment and how the learner results should be interpreted. Individual learner results should be provided as well aggregated scores. Exceptional learners should be identified as well as learners who may require additional attention. As far as possible visual representations in the form of graphs should be given, possible reason for poor performance is given as well as key areas where learners had difficulty. The report should also include attitudinal data as to the problems learners are experiencing at school, views towards the school and classes.

In conclusion, it is clear that exploring International systems has promise for South Africa. Not only can the instruments be used when adequately adapted but also lessons learnt from the development and implementation of such systems can be used in order to obtain a monitoring system that is fair, valid and reliability.

References

Anastasi, A., & Urbina, S. (1997) *Psychological testing* (7th ed). New Jersey: Prentice Hall.

Botha, R. J. (2002) Outcomes-based education and educational reform in South Africa. *International Journal of Leadership in Education 5*(4), 361-371.

Creswell, J. W. (2003) *Research design: Qualitative, quantitative and mixed method approaches* (2nd ed). London: Sage Publications.

Coolican, H. (1999) *Research methods and statistics in Psychology* (3rd Ed). London: Hodder & Stoughton.

Curriculum, Evaluation and Management Centre (CEM) (2002a) Introduction to the MidYIS Project. Retrieved 17 May, 2002, from http://midyis.cem.dur.ac.za.uk/default.asp.

Curriculum, Evaluation and Management Centre (CEM). (2002b) The MidYIS Baseline Tests. Retrieved 3 December, 2002, from http://www.midyisproject.org/thetests.asp

Curriculum, Evaluation and Management Centre (CEM). (2002c) Reliabilities. Retrieved 3 December, 2002, from http://www.midyisproject.org/reliabilities.asp

Curriculum, Evaluation and Management Centre (CEM). (2006) PIPS Reliabilities. Retrieved 5 May, 2006, from http://www.cemcentre.org/

Fitz-Gibbon, C. T. (1992) 'Empower and monitor: The EM algorithm for the creation of effective schools'. In J. Bashi & Z. Sass (Eds), School effectiveness and improvement. Proceedings of the Third International Congress for School Effectiveness. Jerusalem: Magnes Press.

Fitz-Gibbon, C. T. (1996) *Monitoring education: Indicators, quality and effectiveness*. London: Continuum.

Gay, L. R., & Airasian, P. (2003) *Educational Research: Competencies for analysis and application* (7th ed). New Jersey: Merrill Prentice Hall.

Gauteng Department of Education (2005) Gauteng online to reach 1000 Gauteng Public Schools. Retrieved 18/01/20006 from http://www.education.gpg.gov.za/publications/gauteng%20online%20to%20reach%20public%20schools.htm

Greene, J. C. (2005) Mixing methods in evaluation workshop notes. Centurion Country Club, Gauteng. 23-24 June 2005.

Gronlund, N. E. (1998) *Assessment of student achievement* (6th ed). Boston: Allyn and Bacon.

Hager, K.D., & Slocum, T.A. (2005) Using alternative assessment to improve educational outcomes. In *Rural Special Education Quarterly 24*(1), 54-59.

Hendriks, M. A., Doolaard, S., & Bosker, R. J. (2001) School self-evaluation in the Netherlands: Development of the ZEBO-instrumentation. *Prospects XXXI* (4), 503 – 518.

Hendriks, M. A., Doolaard, S., & Bosker, R. J. (2002) Using school effectiveness as a knowledge base for self-evaluation in Ditch schools: the ZEBO-project. In A. J. Visscher & R Coe (Eds.), *School improvement through performance feedback*. Lisse: Swets & Zeitlinger Publishers.

Kline, P. (1993) *The handbook of psychological testing*. Routledge: London.

Linn, R. L., & Gronlund, N. E. (2000) *Measurement and assessment in teaching* (8th ed). New Jersey: Prentice Hall.

Lockheed, M.E. (1996) International context for assessments. In P Murphy, V Greaney, M.E. Lockheed and C Rojas (Eds), *National Assessments: Testing the system*, pp 9-20. Washington DC: The World Bank.

Murphy, K. R., & Davidshofer, C. O. (1994) *Psychological testing: Principles and applications*. New Jersey: Prentice-Hall International.

National Center for Research on Evaluation, Standards and Student Testing. (1999) CRESST Assessment Glossary. Retrieved, June 5, 2004, from http://cresst96.cse.ucla.edu/CRESST/pages/glossary.htm.

National Department of Education. (2002a) National Curriculum Statement - Grades R-9: Languages - English Home Language. Retrieved 17 August, 2005 from: http://www.education.gov.za/mainDocument.asp?src=docu&xsrc=poli

National Department of Education. (2002b) National Curriculum Statement - Grades R-9: Mathematics. Retrieved 17 August, 2005 from: http://www.education.gov.za/mainDocument.asp?src=docu&xsrc=poli

Neuman, W.L. (1997) *Social research methods: Qualitative and quantitative approaches*. Boston: Allyn and Bacon.

Owen, K., & Taljaard, J.J. (1996) *Handbook for the use of psychological and scholastic tests of the HSRC*. Pretoria: Human Science Research Council.

Plomp, T. P. (2004) Quality assurance in Netherlands education. Workshop held at the University of Pretoria, 5 August.

Raffan, J., & Ruthen, K. (2003) Monitoring and Assessment. In J Beck & M Earl (Eds), *Key issues in secondary education: Introductory readings*, pp 28-40. New York: Continuum.

Rowe, K. J. (1999) 'Assessment, performance indicators, league tables, value-added measures and school effectiveness? Consider the issues and let's get real!' Paper presented at Australian Association for Research in Education. Retrieved July 5, 2005, from http://www.aare.edu.au/99pap/row99656.htm.

Rowe, K. J., Turner, R., & Lane, K. (2002) Performance feedback to schools of students' year 12 assessments. The VCE data project. In A. J. Visscher & R Coe (Eds.), *School improvement through performance feedback*. Lisse: Swets & Zeitlinger Publishers.

Ruiz, B., Fitz, P.A, Lewis, C. & Reidy. C. (1995) Computer-adaptive testing: a new breed of assessment. In *Journal of the American Dietetic Association, 95*(11), 1326-1328.

Safer, N., & Fleischman, S. (2005) 'How student progress monitoring improves instruction'. In *Educational Leadership*/February 2005.

Sammons, P. (1999) *School effectiveness: Coming of age in the twenty-first century*. Lisse: Swets & Zeitlinger Publishers

Sanders, W. L., Wright, S. P., Ross, S. M., & Wang, L. W. (2000) Value-added achievement results for three cohorts of roots and wings schools in Memphis: 1995-1999 outcomes. Retrieved July 20, 2004, from: http://www3.sas.com/govedu/edu/research.html.

Sanders, W. L., & Horn, S. (2003) An overview of the Tennessee Value – Added Assessment System (TVAAS): with answers to frequently asked questions. Retrieved August 22, 2003, from http://www.mdk12.org/instruction/ensure/tva/tva_1.html.

Scheerens, J., & Hendricks, M. (2002) School self-evaluation in the Netherlands. In D. Nevo (Ed), School-based evaluation: An international perspective. Amsterdam: JAI Elsevier Science.

Scheerens, J., Glas, C., & Thomas, S.M. (2003) *Educational evaluation, assessment and monitoring: A systemic approach*. Lisse: Swets & Zeitlinger Publishers.

Suen, H. K. (1990) *Principle of test theories*. New Jersey: Lawrence Erlbaum Associates.

Teddlie, C., Kochan, S., & Taylor, D. (2002) The ABC+ model for school diagnosis, feedback and improvement. In A. J. Visscher & R Coe (Eds.), *School improvement through performance feedback*. Lisse: Swets & Zeitlinger Publishers.

Teddlie, C., & Tashakkori, A. (2003) Major issues and controversies in the use of mixed methods in the social and behavioural science. In A. Tashakkori & C. Teddlie (Eds), *The Handbook of mixed methods in the social and behavioural research*. London: Sage Publications.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed). New Jersey: Prentice Hall.

Tymms, P. & Wylde, M. (2003) Baseline assessment and monitoring in primary schools. Paper presented at the Symposium Connectable Processes in Elementary and Primary Section – Bamberg April 2003

Urbina, S. (2004). *Essentials of psychological testing*. New Jersey: John Wiley and Sons, Inc.

United Nations Educational Scientific and Cultural Organisation (UNESCO). (2005). Education for all: Global monitoring report. Retrieved 3 June 2005 from http://www.unesco.org/education/efa.

Ward, L., Hattie, J.A., & Brown, G.T. (2003, June). The evaluation of asTTle in schools: The power of professional development. asTTle Technical Report, #35, University of Auckland/Ministry of Education.

Williams, K. (1999). Mixed quantitative and qualitative evaluation tools: A pragmatic approach. Retrieved March 17, 2003 from: www.cemcentre.org/ebeuk/papers2/Williams%20Kevin.doc.

International Assessment Standards versus local curriculum objectives

International assessment standards and local academic achievement

Ngoato Takalo, North-West University

Introduction

Let me firstly thank the organizers of this conference for inviting me, and also extend a word of appreciation and gratuity to Umalusi for having the confidence in me that I can make a small contribution to the debates. More importantly, from a personal point of view, I see this participation very important because seldom do higher education and high school levels of education come together to share experience, let alone in issues of assessment. I have a strong belief that this will be the beginning of a long lasting partnership.

My professional background is that of a teacher, while my academic background is that of a curriculum worker. I have spent a large part of my career in management and administration. I am providing you with this information in order to prepare you for the kind of paper you will be exposed to. Firstly, it is not an academic paper that is ready for publication. In my daily work I do not have the privilege to really sit and do a library search and write a proper academic paper. Secondly, bear with me if this paper has a bias towards policy. That is what I do every day. I shall attempt to bring together the issues of curriculum and assessment in a simple manner that makes the most sense to me, and hope you will share it with me.

International assessment standards vs local curriculum objectives

In this paper I am just going to ask questions and raise issues that relate to what one could refer to as an obsession with international standards and whether we really should be pre-occupied with that, or whether we should use those standards as a reference point when we ensure that learners have gained sufficient knowledge and achieved the objectives as according to local curricula.

Educational standards provide criteria for learning and ensure quality in educational learning programmes. Assessment cannot be designed without agreement about what children should learn, which is the content of the curriculum. Learning, after all, is the heart and soul of education. Educational learning programmes are a product of a curriculum that is informed by a particular world-view. Curriculum frameworks among countries are not the same.

The issue of a curriculum philosophy for a country, which carries the ethos of that country, is a critical matter. I believe that a curriculum should instill in its consumers a national pride. A curriculum needs to seek a balance between the needs of individual students and the requirements of the society and the economy. It needs to foster the development of a workforce which is highly skilled and adaptable with an international perspective. There should not be a disjuncture between the philosophy and national achievement objectives, to monitor student progress against these objectives as well as barriers that might exist. How many of our curricula contextualise the national attitudes and values in the school curriculum? Do we encourage, through our policies, learners to learn self management and work and study?

If certain countries depend on others for assessment standards what happens with the curricula of the dependants because as standards stay steady over time, the curriculum should be up to date; which means regular continuous review and update? Does it mean that whether it suits them or not, the dependent countries must review their curricula at the pace of metropolitan countries? If this is the case, then the stability of the achievement of national curricula objectives becomes questionable. Surely, countries decide on what their core-curricula should be, and how assessment of achievement in that subject should be carried out. These core curricula might differ from one country to the other. That is why the standards would be unique from one country to the other. There are voices in those countries, like the US, who do not have a national assessment system, calling for a national standardised assessment system. This according to Ravitch (1993) supporting Lamar Alexander, who was then the US secretary of education, would be in support for national education goals.

There should not develop a perception that high academic standards are discouraged. That would be suicidal. High standards for all students are important and encouraged if they fit the objectives and expectation of a particular country. Standards and assessments are critical because, in actual fact, they should allow teachers, students, parents and communities to determine whether they are aiming high enough and whether they are making progress.

Curriculum, be it a state-, a provincial- or national- comes before assessment. Its goals and objectives set the pace for standards and quality of assessment. Therefore, those who conceptualise curricula, turn those concepts into policy, curriculum developers interpret those into teaching/learning material, cannot work separately from those who develop assessment policies and instruments. This is all about what learners should learn and what countries should achieve. This, in turn, means that staff development should help teachers to understand linkages in the chain of their work and the pride they need to display in studying and comparing international standards in order to set their own.

The issue of adopting foreign standards of assessment is a historical one in Africa and other formerly colonised countries. Metropolitan educational models tended to influence local ones even after independence. De-colonised countries clung to the metropolitan models in fear of "falling" standards and relegation to second class international citizenship. During those days nothing local was good enough. Countries such as Tanzania were the first ones to see positive steps to adopt more truly indigenous national systems of education. This would lead to a situation where local and environmental factors are taken into consideration, such as social, economic and political situations of the countries and for those former colonies that have opted for democracy, in a democracy every person has the right and the responsibility to participate in the control of public institutions. This means that citizens, parents and public officials have the right to know how well schools are fulfilling their responsibilities. That is why local but internationally accepted standards is a better option.

In conclusion, I would like to leave you with some of the things that I wish could happen:

1. Higher education institutions have been reluctant to enter into the standards discussion and seem content to remain at arms-length. My wish is that there could be more formal discussions between these two levels of education for the benefit of our learners and societies.
2. I wish higher education institutions, with their complex curricula, and understandably so, could show some ability to be adaptable by showing a little inclination to revamp their curricula to align with those of high schools. This would, perhaps, stop the ranking of high schools by higher education institutions with some schools being disregarded.

Many forces are converging, though, to enforce a stronger nexus between high school and higher education. Should this initiative succeed, adopting international standards slavishly will come to an end because higher education institutions are autonomous and set their own standards.

As you are gathered here and engaging with this very important topic some people out there may be thinking that you really are not utilizing your time effectively; you should be doing something better out there than talking about how to raise the achievement levels. These are the people who may have forgotten where they come from. I can assure you that we, in higher education, welcome these types of conferences because they assure the quality of our prospective students. The only thing is, we also need to jump onto the bandwagon.

Thank you and I wish you all of the best in your further deliberations.

Constructing an integrated model

For the quality assurance of school-based assessment in preparation for the National Senior Certificate

Rufus Poliah, South African National Department of Education

Abstract

School-based assessment constitutes part of the assessment that leads to the issue of the exit certificate, after 12 years of schooling in South Africa. The high stakes nature of this certificate warrants a high level of reliability and validity of the assessment. There has been much debate about the reliability of the school-based assessment component, given the challenges associated with the implementation of school-based assessment in any large system. Currently, the system is mainly dependent on a model of statistical moderation that brings the school-based marks within an acceptable deviation of the examination mark.

This paper explores the challenges relating to the quality assurance of school-based assessment in South Africa and presents a model for the quality assurance of school-based assessment that is built on three fundamental pillars, i.e. (a) setting of clearly defined standards; (b) providing appropriate support and guidance; and (c) monitoring and evaluation of the assessment process and the outcome. The setting of standards incorporates input, process and output standards and the support and guidance relates to professional and administrative support required for implementation. In the evaluation of the final outcome, this model argues for the integration of statistical and qualitative data that will ensure that every mark is a reliable indicator of the performance of the learner. This model could be progressively implemented as the new curriculum is phased in, commencing in 2006, and, if successfully implemented, will go a long way in improving public confidence in the reliability of school-based assessment.

Introduction

The international trend in school assessment is towards a combination of both external assessment and internal assessment. Internal or school based assessment is seen as a tool for better learning and an integral part of the teaching and learning process, where equal importance is attached to diagnostic and formative; as well as to summative assessments. However, with the introduction of internal assessment or school-based assessment as part of

the final certification process, the reliability of this assessment component, is always in question. The usefulness of this form of assessment is always considered against the background of its lower reliability. Therefore a model that seeks to address the quality and standard of this assessment needs to be vigorously pursued. Statistical approaches to quality assurance are restrictive, but are adopted by most assessment bodies because they are easy to implement. Therefore, this paper provides an alternative that seeks to integrate both statistical and qualitative modes of quality assurance.

What is school-based assessment?

Izard (2001) and Raivoce and Pongi (2001) indicate that school-based assessment is often put in place to collect evidence on what students have achieved, especially in important learning outcomes that do not easily lend themselves to pen and paper tests. Daugherty (1994) further explains that this type of assessment is used because of the gains in validity, which arise from student tasks that can be assessed in a wider range of contexts and more frequently than is possible within the time-limited constraints of a controlled written examination.

School-based assessment (SBA) can be more broadly defined as assessment that is conducted by the teacher in the classroom. School-based assessment is sometimes referred to as internal assessment, coursework or continuous assessment (CASS) and includes, *inter alia*, the following types of assessment:

1. Assessment tasks are designed, administered and judged by the teacher
2. Assessment tasks are centrally designed, but administered and judged by the teacher

SBA is conducted as the learning process takes place and is used to influence or inform the learning process. It comprises a variety of assessment methods that may be formal and informal. The following assessment methods and instruments have been identified as components of SBA, i.e. assignments, oral tests, case studies, practical exercises, projects, portfolios, personal interviews, questionnaires, role-play activities and simulations. SBA refers to any of these forms of assessment conducted internally which are collated for summative judgement purposes in order to feed into the exit point assessment.

Lusby (2004) has explained that for many, internal assessment has been solely an assessment of those skills that the external examination does not cover, but it is now being accepted more widely that internal assessment for qualifications can include an even wider range of learning outcomes, even those traditionally assessed by an external examination. Internal assessment, therefore, allows for assessment to take place at the time of learning and to be integrated with learning. If learning outcomes are only assessed in an external examination at the end of the academic year, then memorisation also becomes one of the outcomes that are being tested.

Why school-based assessment

The arguments in support of school-based assessment are varied, but one of the reasons is that SBA extends the scope of the assessment to include not only formal written work but also oral and practical work and in some cases, personal attributes. This type of assessment

also provides for assessment that will be more diagnostic and detailed, increasingly cumulative and integrated with the learning process and culminating in a terminal evaluation.

In the days when examinations dominated the assessment arena, examination systems placed a lot of importance on the issues of reliability and comparability of results. Broadfoot (1995) explains that in a high stakes environment where test results determined life chances, there was an emphasis on reliability so that the assessment was seen to operate fairly and consistently. Reliability was then of more importance and validity – whether the test does measure what it is intended to measure – was often of subordinate importance.

However, the situation has started to change over the last two decades (Grima, 2003), mainly because as Gipps (1999) explains:

> ….the focus has shifted towards a broader assessment of learning, enhancement of learning for the individual, engagement with the student during assessment, and involvement of teachers in the assessment process.

According to Grima (2003), the rise of school-based assessment is a result of this change.

There are a number of advantages when classroom-based assessment is included in external examinations. Taylor (2003) summarises these advantages into four useful categories viz:

1. Evidence that teachers are preparing students to meet mandated curriculum and performance standards (opportunity to learn);
2. Broader evidence about student achievement than what can be obtained from examinations administered in a brief period;
3. Opportunities to assess knowledge and skills that are difficult to assess via standardised large-scale tests/examinations; and
4. Opportunity to include work that more closely represents the real contexts in which knowledge and skills are applied.

Challenges associated with the implementation of CASS

There are several problems that are associated with the implementation of classroom-based assessment and the following are some of the main problems:

1. The level of teacher preparation for school-based assessment is not adequate in most cases. According to Schafer and Lissitz (1987) and Stiggins and Bridgeford (1986), few teacher preparation programmes provide adequate training for the wide array of assessment strategies used by teachers. According to Taylor (2003), teachers must be taught how to select, modify and develop assessment tasks, as well as how to consistently evaluate the work of learners and how to develop scoring instruments, before school-based assessment can be implemented. She argues that a significant professional development effort is required, even for teachers who have received adequate training in assessment methods.
2. There is always a variation in the scoring of assessment tasks among teachers, particularly when the assessment tasks are not the same and when each individual teacher develops the assessment rubric or the marking guideline. This is supported by Taylor (2003), who

concluded, based on recent research, that a critical problem when including classroom-based assessment in large-scale test programmes, is the degree to which raters agree with one another when evaluating student work.

3. The implementation of CASS leads to an increase in both teacher and learner workload, i.e. extensive record keeping and monitoring of individual learners are required.
4. Variance in the final assessment can result from learners having been assisted by their teachers, peers or parents, or the differences in the difficulty of classroom-based tasks and the degree to which the classroom-based task represents the domain of reference.
5. Different schools may adopt different approaches to continuous assessment and this may impact on comparisons across learners.

It is these challenges that make the quality assurance of school-based assessment a critical issue. As Maxwell (2003) explicates, it is necessary to develop consistency in teacher judgement of student achievement and to ensure public confidence.

The current status of CASS in the South African context

Quality education is a national priority in South Africa. Emerging from an apartheid education, the new government has embarked on the overhauling of the entire curriculum. The National Curriculum Statements (NCS) for the General Education and Training band (Grades R–9) is currently being implemented across all grades in this band, while the National Curriculum Statements for the Further Education and Training band (Grades 10-12) are being phased in at Grade 10 level in 2006.

As from 2008 a new certificate, the National Senior Certificate (NSC), will be awarded for the achievement of the exit-level learning outcomes stipulated in the National Curriculum Statements (Grades 10-12). For Grades 10, 11 and 12, learners will be assessed internally according to the requirements as specified in the Subject Assessment Guidelines. The internal assessment mark allocated to assessment tasks completed during the school year will be 25% and the end- of-the-year assessment mark, 75% of the total mark.

School-based assessment or internal assessment is also a part of the current curriculum and at Grades 10-12 level it is a component of the final assessment that leads to the issue of the current certificate, the Senior Certificate (SC). The final written examination constitutes 75% and the internal assessment constitutes 25% of the final mark.

The South African education system is a national system, with the national Department of Education responsible for policy development and ensuring policy compliance. The nine provincial education departments are responsible for the implementation of policy. Each of the nine provincial education departments, as part of their education provision function, is responsible for the implementation of examinations and school-based assessment in the schools under their jurisdiction. In order to ensure quality in the assessment leading to the exit certification, Umalusi, the Quality Assurance Council for General and Further Education and Training, has been established to take final responsibility for the standard and quality of assessment leading to the issue of the exit certificates.

Despite the challenges confronting the implementation of SBA in South Africa, which are similar to those experienced in other parts of the world, the compulsory inclusion of SBA as part of the final assessment leading to the Senior Certificate, was made mandatory by the then Minister of Education, in 2001. This was a bold step given the disparities in the system and the capacity of teachers to implement this new approach. The overriding motivation for the inclusion of SBA was the need to use continuous assessment to promote the culture of teaching and learning in schools. There was concern about the reliability of the marks awarded to learners at schools and given the size of the system it was not possible to establish effective internal moderation systems at provincial level. Therefore, Umalusi adopted a model of statistical moderation where the SBA marks are adjusted within a certain range of the adjusted examination marks. Reflecting on the last five years of SBA implementation at Grade 12 level, it can be concluded that the decision made in 2001, has moved the system forward.

In order to ensure that every mark submitted by a school is a reliable and valid indication of the performance of the learner, Umalusi carries out a statistical moderation of the SBA marks. Statistical moderation of SBA is undertaken per institution, per subject. The mean and standard deviation of the examination mark (written paper) is used in this process. The "mean" being the average score of the institution in that subject and the "standard deviation" is an indication of the spread of the scores in that subject at the particular institution.

After the examination scores have been standardised, the mean of the examination score of a particular subject at a particular centre is compared to the mean of the SBA score. If the mean of the SBA is within a certain range of the examination mean, then the SBA mean is accepted as is. If the mean of the SBA score is either too low or too high, it is brought to within a certain range of the examination mean. The standard deviation of the examination mark for that centre is also compared to the standard deviation of the SBA. The standard deviation of the SBA is adjusted to the same standard deviation as that of the examination mark.

The adjusted CASS mark of the learner is then added to the adjusted examination mark in the ratio of 25:75, i.e. CASS constitutes 25% of the final mark of the learner.

Each provincial education department is responsible for the internal moderation of CASS. The effectiveness of the provincial moderation systems varies from one province to the other. Moderation is supposed to take place at the three tiers of the provincial education department, i.e. school level by the senior teachers or heads of department, cluster level, where schools get together to support and moderate samples of assessment tasks and at provincial level, where samples of portfolios are collected from schools and moderated at a central venue, or schools are visited for this purpose. In its 2005 report to the Minister of Education, Umalusi, the Quality Assurance Council, indicated that problems regarding the reliability of CASS continue to dog continuous assessment, but Umalusi's statistical moderation of CASS will enhance confidence in the scores that accrue from this process.

It is in this context that a model for the quality assurance of SBA is proposed that shifts the dependence from statistical modes of moderation to a mixed mode of qualitative and statistical mechanisms.

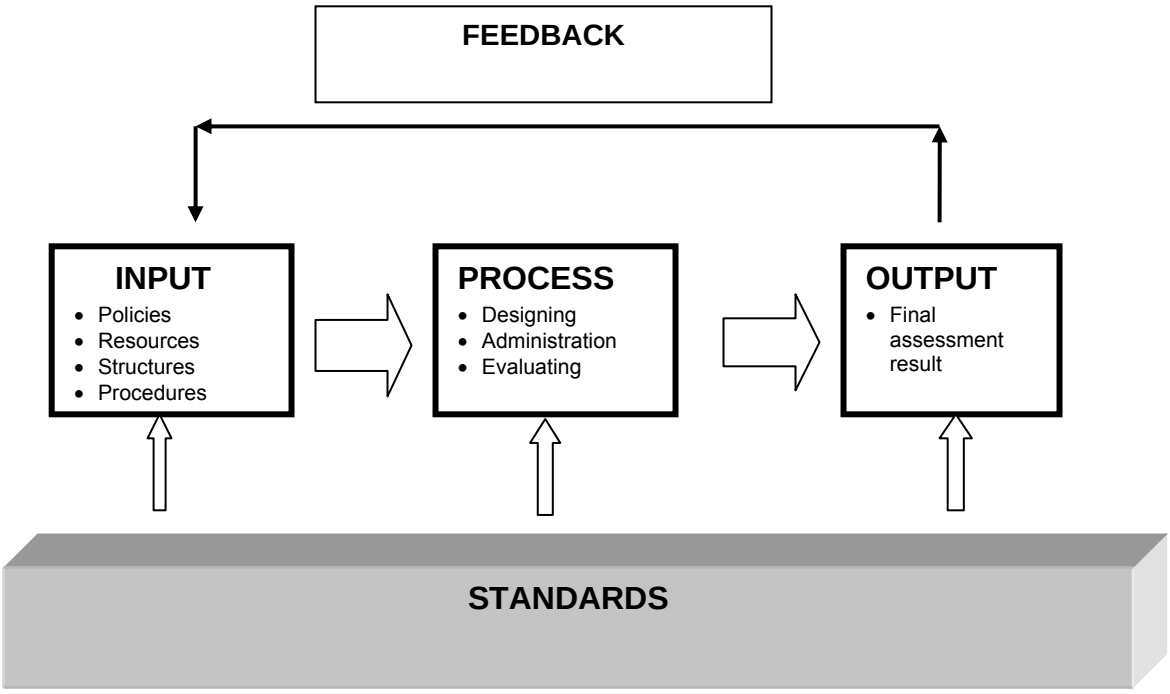
A model for quality assurance of CASS

The model for quality assurance proposed for CASS in South Africa is based on the challenges confronting the system and an understanding that quality assurance must include a focus on the quality of each component of the assessment system, and not just the end product. This model takes as its point of departure the establishment of clearly defined standards that are accepted and understood by all participants in the assessment process, followed by appropriate support, training and guidance to the implementers of assessment and the final monitoring and evaluation with a view to providing essential feedback to the system so that the improvement cycle can be initiated. This model also acknowledges the size of the system, the differences in terms of resource provision, the urban/rural divide and the workload of teachers.

Conceptual framework for quality assurance of CASS

The conceptual framework for the quality assurance of CASS is based on the Input, Process, Output (IPO) Model. According to the IPO model there is a causal link between the inputs, processes and outputs. This means that the variables at input level have a direct influence on the variables at process level, which in turn influence the variables at output level. CASS as an assessment system has clear inputs, processes and outputs. It is vital that the critical inputs, processes and outputs are identified and quality assurance will focus on each of these components.

Diagrammatic representation of the IPO model



The following is a summary of the inputs, process and outputs relating to CASS.

The inputs include the elements of the assessment system which mainly relate to the following:

- Policies, regulations, guidelines, circulars developed for assessment and moderation
- Physical and human resources
- Educators and internal moderators
- Competency of educators and internal moderators
- Structures and procedures for assessment at the different levels in the system
- Implementation plans
- Development programmes for educators and internal moderators
- Production and security of assessment instruments
- Monitoring systems
- Recording and reporting systems
- Evaluation and review processes

The assessment process will cover all aspects relating to the implementation of the assessment process and these include the following:

- Designing the assessment instrument
- Internal moderation of the assessment instrument
- Administration of the assessment instruments
- Evaluating the assessment evidence
- Recording and reporting of the assessment outcome

The outcome relates to the final assessment result that emanates from the assessment process. The final assessment result must be reliable, valid and enjoy public confidence.

Implementing the model

Setting of standards

Umalusi and the Department of Education (DoE), have the joint responsibility to provide standards for good assessment. Standards for assessment must relate to the input, process and output of assessment. The following are the standards that must be clearly articulated to ensure valid and reliable assessment:

Curriculum Standards

Curriculum standards are contained in the National Curriculum Statements (NCS) and they are articulated as learning outcomes, assessment standards and the prescribed content. They provide the educator with a description of the knowledge, skills, attitudes and values that must be demonstrated at the relevant grade, within the subject learning area. The assessment standards in the NCS are associated with the learning outcomes and they provide the

educator with a description of the scope and range of the performance for each learning outcome. They also serve as a benchmark to measure whether the learning outcome has been achieved or not.

In the South African curriculum, the curriculum standards are well defined, but for the purpose of SBA, it is proposed that the curriculum standards differentiate between the knowledge and skills that should be measured through a pencil and paper examination and those that are best measured through other forms of assessment that are administered at school by the educator. This distinction does not imply that SBA should exclude knowledge and skills that could be measured through pencil and paper assessment, but it will allow for a distinction in the standardisation of the final assessment outcome, allowing for related assessment constructs to be standardised and correlated.

As South Africa moves towards the implementation of the new curriculum statements, learning and teaching support material (LTSM) is being developed that will help educators to fully grasp the essence and requirements of the curriculum standards.

Performance standards

Performance standards assist educators in measuring how well learners have performed with reference to a specific curriculum standard. Performance standards may take the form of exemplars of learner performance, indicating clearly the grade awarded to the performance and the reasons for the award. These standards contribute to establishing a common understanding among assessors and ensure consistency in the awarding of grades and scores.

Assessment system standards

An assessment system comprises a description of the policies, regulations, structures and procedures that make up assessment. The components of the assessment system are listed above as inputs. The assessment system standards must also describe the processes and procedures for the implementation of assessment and this should include, *inter alia:*

- Process of design and production of assessment instruments;
- Conduct of assessment;
- Awarding of marks/scores; and
- Evaluation and review of assessment practice.

Moderation Standards

Moderation is the process of ensuring that assessments are conducted in accordance with agreed practices and standards. Minimum requirements for internal and external moderation systems that must be adhered to by all providers of assessment must be stipulated for the moderation process.

Support and Guidance

Support and guidance refer to the various kinds of support and guidance provided to all persons engaged in the implementation of SBA at all levels in the system. For the purpose of this model, the following three key categories of persons engaged in the implementation of CASS will be discussed:

1. Managers

2. Subject advisory support
3. Educators

Support and guidance to managers

Managers, in this context, will include officials responsible for the management of SBA at national, provincial, district and school levels. These individuals are responsible for the development of policies, guidelines, circulars, drafting implementation plans, setting up processes and procedures, making available resources, training, etc. These managers are responsible for ensuring that the assessment system standards are complied with. In most cases these managers are educators who were teachers, but have had little or no management training or experience. It is therefore essential that, as a first step, all managers be provided with intensive management training relating specifically to management of assessment. In order to ensure support for a new policy, managers need to be exposed to the policy first before the implementers of the policy are trained. This has been an oversight in the South African situation where, due to time constraints, managers (e.g. school principals, district managers, etc.) have been omitted from the training relating to the new curriculum policy, but they are nonetheless expected to manage policy implementation at the level they are operating. Unless there is "buy in" and support from managers at strategic levels, policy innovation is not guaranteed. The training of managers is twofold, i.e. training in the contents of the policy and training relating to the management of the policy.

Support and guidance to subject advisors

Subject advisors are the arbitrators and mediators of policy, which implies that successful policy implementation is to a large degree dependent on the effectiveness of subject advisory services. In South Africa there is a serious shortage of subject advisors, for example in one province where there are 1 000 schools offering Grade 12, there are only 3 Mathematics subject advisors (Singh, T, 2004). In order to ensure effective subject advisory support, provincial education departments must appoint the required number of subject advisors and the following support to subject advisors is essential:

- Training in the policy and procedures relating to SBA
- Training in the content areas of the subject
- Training in the forms of assessment relating to SBA
- Training in their subject advisory role
- Appropriate support material to assist them in the execution of their duties

Subject advisors must be experts in their fields to command the respect of the teachers they support. The assumption that subject advisors, by virtue of their positions, are the leaders in their fields is a misnomer.

Support and guidance to educators

Educators are the key implementers in the process of SBA and the success of implementation is therefore determined by their competency in this area and their commitment to this mode of assessment. Educators must be thoroughly trained, both at pre-service and in-service levels. At pre-service level the Department of Education will ensure that all institutions involved in teacher development will focus on learner assessment. For practising educators it is proposed that they all be trained using an accredited assessment course. The assessment course must be customised to focus specifically on the design of

assessment instruments and the evaluation of assessment evidence. The South African Qualifications Authority must accredit educators that are trained and who have successfully completed the assessment course. Educators that are accredited must be further subjected to compulsory in-service training that is scheduled at least once in three years.

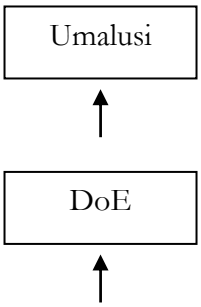
To further support educators in the implementation of SBA, it is also proposed that the Department of Education embark on the development of item banks, focussing specifically on items that can be used for SBA. Items can be generated by inviting educators to develop items, based on specifications provided, or alternatively, competent educators can be selected to develop items. These items must be subjected to review and moderation before they are placed in the item bank. Each item will be accompanied by marking rubrics or a marking memorandum, as the assessment task demands. The item bank can be made available to educators both electronically and in hard copy format. Teachers can then use these items that are designed and approved nationally and this will allow for improved reliability in the assessment conducted by teachers.

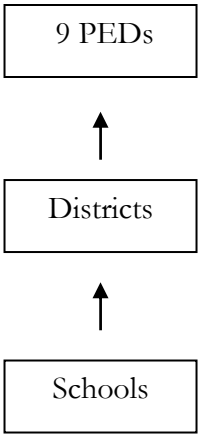
In addition to item banks, teachers can be successfully supported by being provided with exemplars of learner evidence, that is graded in accordance with the assessment rubrics or marking memorandum. The Department of Education piloted this initiative with examination related question in History, in 2005 and it is proposed, as part of this model, that exemplars and their scoring should be extended to assessment tasks that are used during SBA.

The ongoing support and guidance provided by subject advisors, principals and heads of department must be maintained at the most effective levels and cannot be compromised. It is the responsibility of the provincial education department to ensure that subject advisory support as well as management support at school level, are effectively implemented and maintained

Monitoring and Evaluation

Monitoring and evaluation, which is the final leg in the quality assurance process, is the responsibility of the provincial education departments (PED), the national Department of Education (DoE) and the Quality Assurance Council (Umalusi). Each of these institutions has a responsibility to monitor and evaluate at a particular level. The PED must take responsibility for the schools and the districts in that province. The DoE will exercise national oversight over all PEDs and Umalusi will verify the findings of the DoE. This can be illustrated as follows:





Monitoring and evaluation by all three key role-players will address the following three components:

1. Evaluation of the assessment system (input)
2. Monitoring of the implementation of the assessment process (process)
3. Moderation of the assessment outcome (output)

Evaluation of the assessment system (input)

The assessment system includes all the inputs, i.e. policies, regulations, structures, processes and tools used in SBA, as articulated in paragraph (6.1). The evaluation process will commence with each PED taking responsibility for evaluating each school under its control.

Each school will be requested to complete a self-evaluation instrument, which must be submitted together with the required supporting evidence to the district office. The district office will evaluate the evidence and conduct a verification visit to each school under its control. The district will then provide a detailed report to each school on its strengths and weaknesses, with recommendations for improvement. Schools, on the basis of the report, will develop an improvement plan, in conjunction with the district, clearly stating the time frames for implementation. It is the responsibility of the district to monitor the implementation of the improvement plan. The district will submit a detailed report on each school to the provincial head office.

The provincial head office will conduct sample visits to the schools within the district to verify the contents of the report. The PED will also evaluate the district and its support provided to the schools. The PED will compile a composite report on the schools and districts under its jurisdiction and this will be submitted to the DoE. The DoE will conduct sample visits to selected schools and districts in the province to ensure that the evaluation process has been accurately complied with. The DoE, on completion of its evaluation, will report accordingly to Umalusi and request Umalusi to evaluate the assessment bodies for compliance with the requirements of the assessment systems.

The report from Umalusi, the Council for Quality Assurance in South Africa, will serve as the final report on the status of the assessment systems of the nine provincial education

departments. This report will articulate the areas where improvement is required, and the DoE will take responsibility for ensuring that these improvements are effected according to the stated time frames.

Monitoring the implementation of the assessment process

Monitoring of the implementation of SBA will be conducted by:

- The senior management team at school level
- District officials/subject advisors for schools
- Cluster leaders for schools in the cluster
- Provincial officials for schools in the district
- National Department officials for schools in the province

Monitoring will focus on the actual implementation of the assessment process and will cover the following aspects:

- Ensuring that SBA is implemented at schools
- Ensuring compliance with policy
- Assessment tasks are designed in accordance with the subject assessment guideline;
- Assessment tasks are administered in accordance with the principles of good assessment
- Assessment evidence is marked according to the marking guidelines
- Marks are totalled and recorded as per policy

Reports on the monitoring at each of these schools must be completed and submitted to the next level.

Moderation of the assessment outcome (output)

The IPO model removes the focus of moderation from the outcome and places it on all three components, i.e. the input, process and output. This model also pre-supposes that if the input and process have been thoroughly quality assured, the outcome should demonstrate a higher reliability.

For the moderation of the final assessment results, it is proposed that the assessment tasks that mirror the examinations should be separated from the assessment tasks that measure skills that cannot be measured through pencil and paper tests. Assessment tasks that mirror the examinations, measure the same assessment construct as the examinations and therefore they can be statistically moderated to the examinations, using the current approach. The current formula adjusts the SBA marks to the examination marks, allowing for a 5–10% tolerance range in the mean. The tolerance range of 5–10% allows for a better performance in the SBA, since learners are tested on a smaller volume of work and the assessment is administered immediately after the teaching and learning.

In the case of assessment tasks that measure skills that do not mirror the examination, it is proposed that the assessment tasks and the assessment evidence presented by the candidates should be moderated at the four tiers, viz:

- School level by the head of department or the senior subject teacher

- Cluster level or district by the cluster group or subject advisors
- Provincial level by moderators appointed by the PED
- National level by DoE and Umalusi

The cluster, which includes a group of schools, must be chaired by an appointed cluster leader. The responsibility of the cluster leader and the cluster team will be to moderate samples of portfolios from each of the schools in the cluster and to ensure common standards across the schools in the cluster. The subject advisor will exercise overall control over all the clusters in his/her district.

At provincial level, provincial moderators must be appointed who will moderate samples of portfolios from the clusters. The focus of this moderation is not the individual schools, but the cluster as a whole. This tier of moderation will moderate assessment tasks and the assessment evidence to ensure that the standard of the clusters across the whole province are consistent with the curriculum standards. Provincial moderation can be scheduled at the end of the assessment process and provincial moderators can be appointed following the same procedure as markers, for a specific duration of time. Samples of portfolios from the various clusters will be collected and brought to a central venue, when the moderation takes place.

The DoE will also appoint national moderators who will moderate a sample of portfolios from each of the provinces. The focus of this moderation is to ensure that a consistent standard is applied across all provinces. This level of moderation can be scheduled while provincial moderation is in progress, since the national moderator is an external moderator to the provincial moderation process.

After the national moderation process has been completed, Umalusi, the Quality Assurance Council, will conduct its own verification of the moderation process, relating to this aspect of SBA. Umalusi, based on it own moderation, will decide whether to accept this component of the SBA marks as is, or adjust where necessary. The advantage of separating the SBA into skills; and paper and pencil tasks, is that face moderation can be limited to the skills based tasks and statistical moderation can be applied to the tasks that are paper and pencil, in their focus. This reduces the demand on face moderation

The verification process by Umalusi, will *inter alia*, include a scrutiny of the following reports:
- Evaluation of the assessment systems (inputs)
- Monitoring of the assessment process
- The moderation of the assessment outcome

The important process of feedback must be provided at each level of moderation to ensure that the different levels are informed of their shortcomings and strengths.

Challenges and Implementation Concerns

The challenges relating to this model are as follows:

1. There has to be support for this model by all key roleplayers in the system, since the model advocates quality as an integral part of each process and phase.
2. This system can only be effective if the appropriate human resources are provided at each level of the system. This includes educators, managers and subject advisors.
3. Effective training of educators is pivotal to the successful implementation of SBA and its quality assurance. The Department of Education is currently finalising a Framework for Teacher Development, which will underpin all training in South Africa.
4. The establishment and common understanding of clearly articulated standards across the entire system must be prioritised.
5. The legacy of apartheid has created a wide gap between schools, based on their previous racial profile and their location and therefore any model that seeks to ensure quality across the system must take this as a point of departure.

Conclusion

The model proposed in this paper shows that:

- Statistical moderation does not have to be the sole determinant in the quality assurance of SBA;
- Statistical moderation has a role to play in the correlation and moderation of assessment constructs that are similar;
- The qualitative data from the inputs, process and outputs of assessment will provide useful information in the monitoring and evaluation of the final assessment outcome;
- Clearly defined standards for the key assessment phases should be the starting point in the quality assurance process.

This proposed model does confirm that statistical and qualitative modes of moderation can co-exist in the quality assurance of SBA in the South African context. This model will provide the basis for an engagement on quality and will lead to a nationally accepted model. It is anticipated that over the next two years considerable progress will be made towards improving the reliability of SBA and introducing the NSC, that is credible and has public currency.

References

- Broadfoot, P (1995) Performance Assessment Perspective: International Trends and Current English Experience. In H Torrance (Ed) *Evaluating Authentic Assessment*. London: The Falmer Press.
- Department of Education (2005): The National Certificate : A Qualification at Level 4 on the National Qualifications Framework (NQF)
- Dougherty, L (1992) Quality Assurance, Teacher Assessment and Public Examinations. In W Harlen (Ed) *Enhancing Quality in Assessment*. Newcastle Upon Tyne: Athenaeum Press.
- Gipps, C (1999) 'Socio-cultural; Aspects of Assessment'. *Review of Research in Education,* 24, 355 – 392.
- Grima G (2003) 'School Based Assessment: Implementation Issues and Practices'. Paper presented at the 21st Annual AEAA Conference. Cape Town: South Africa.

- Harlen, W (1994) *Enhancing Quality in Assessment*. London: British Educational Research Association.
- Izard, J (2001) Implementing School-Based Assessment: Some Successful Recent Approaches in Australia and the Philippines. Proceedings of the First international Conference of the Association of Commonwealth Examinations and Accreditation Bodies. Reduit: Mauritius Examinations Syndicate.
- Lolwana P (2000): 'A Standards-Driven Reform in the South African Schools Examinations'. A paper presented at the 18th Conference of the AEAA.
- Lusby S (2004): 'Practical Applications of Moderation in a Regional Qualification'. Paper presented at the ACEAB Conference in Fiji.
- Raivoce, A and Pongi V (2001) 'School Based Assessment: A First Hand Experience in the Small Island States of the South Pacific'. Proceedings of the First international Conference of the Association of Commonwealth Examinations and Accreditation Bodies. Reduit: Mauritius Examinations Syndicate.
- Singh T (2004) School-Based Assessment: the interface between Continuous Assessment (CASS) and the external summative examination at Grade 12 level with special focus on Mathematics and Science. M. Ed Thesis: University of Pretoria.
- Taylor, C S (2003) 'Incorporating Classroom Based Assessment into Large Scale Assessment Programs'. Paper presented at the Annual IAEA Conference. Manchester: United Kingdom.
- Schafer, W D and Lissitz R W (1987) 'Measurement training for School Personnel: Recommendations and reality'. In *Journal of Teacher Education*, 38 (3), 57 - 63
- Stiggins, R J and Bridge ford N J (1986) 'Classroom Assessment: A key to effective education'. In *Educational Measurement: Issues and Practices*, 5(2), 5 – 17.
- Umalusi (2005) Report on the Quality Assurance of the Senior Certificate Examination.
- Umalusi (2004): Policy for the Quality Assurance of Assessment for the GETc and FETC.

Towards piloting school based continuous assessment at middle basic level

The Zambian approach

Newton N. Mutanekelwa and Gabriel G.M. Mweemba, Examinations Council of Zambia

Abstract

Over the years, national school examinations in Zambia have been used for certification and selection of candidates at primary, junior secondary and senior secondary school levels. School-based continuous assessment marks have neither been considered nor used for certification and selection except in some practical subjects.

The 1977 Educational Reforms provided for the use of continuous assessment (CA), but this provision has not been generally applied especially at the Grade 7 level. The Ministry of Education policy document, *Educating Our Future* (1996) stipulates that the Ministry of Education will develop procedures that will enable teachers to standardise their assessment methods and tasks for use as an integral part of the teaching and learning process. The Ministry of Education *Strategic Plan 2003-2007* re-emphasises the need for integrating school-based continuous assessment results with the final examination results for purposes of certification and selection.

This paper discusses the measures Zambia is taking to standardize the school based continuous assessment procedures and the adopted approaches towards piloting of the school based continuous assessment programme. Finally the paper discusses the challenges and the way forward for the programme.

Continuous assessment

Continuous assessment is an on-going, diagnostic, classroom-based process that uses a variety of assessment tools to measure learner performance. It is a formative evaluation measure conducted during the teaching and learning process with the aim of influencing and informing the overall instructional process.

The most important feature of the above definition is that continuous assessment is **a process.** Since it is a process, it must have a product. In order to produce this product there is need to put in place certain necessary inputs.

Background

Over the years, examinations in Zambia have been used for selection and certification. No formal considerations have been placed on the school-based continuous assessment results as a component in the final examinations except for some practical subjects.

The 1977 Educational Reforms explicitly provided for the use of continuous assessment (CA). This provision has not been generally applied especially at the Grade 7 level.

Educating Our Future (1996) and Ministry of education *Strategic Plan 2003-2007* state the need for integrating school-based continuous assessment results with the final examination results for purposes of certification and selection. Furthermore, the policy document stipulates that the Ministry of Education will develop procedures that will enable teachers to standardise their assessment methods and tasks for use as an integral part of school-based continuous assessment.

The current examination system in Zambia consists of final external examinations only. There is no consideration taken with regard to continuous assessment that is done by teachers in schools.

Objectives of continuous assessment

- To provide diagnostic feedback to learners and teachers
- To combine school based assessment results with the final examination result for learner selection and certification.

The continuous assessment programme

In order to set in motion the implementation of the Continuous Assessment programme the Examinations Council of Zambia was tasked to:

- recommend the plan for implementing Continuous Assessment (CA).
- recommend the Training Plan for preparing teachers in implementing the School-Based Continuous Assessment.
- explore ways of ensuring transparency, reliability, validity and comparability in using School-Based Continuous Assessment results.
- agree on Common Assessment Tasks and/or Learning Outcomes to be identified in the syllabuses for School-Based Continuous Assessment.
- consider the development of the Teacher's Manual on Continuous Assessment .
- consider the nature of Summary Forms for the recording of marks that should be provided to schools (recording procedures).

Activities in the continuous assessment programme

Site visits

Site visits to other countries within the sub-region with previous and relevant experience in conducting Continuous Assessment were made. The following countries were visited:

- Malawi
- Namibia
- Tanzania

Training and orientation

In order to set in motion a training and orientation programme for provincial, district, zone and school officials and to produce materials and instruments, a pilot sample of provinces, districts within the provinces and schools within the districts had to be selected. The grade level at which the pilot was to start was Grade 5. This grade level was selected because it is the middle grade in the primary school. Recording of CA marks for integration with the national examination at grade 7 would start at grade 5.

The sample included an urban, peri-urban and rural province, urban and rural district in each province and GRZ, Community and private school representation.

All the Grade 5 teachers, heads of schools, ZIPs, SIPs and guidance teachers in the pilot schools were trained. All DEBS, DESOs, standards officers and statisticians in the pilot districts were oriented on the operations of the CA system. The training and orientation were conducted by a combined team of officers from ECZ, CDC and Directorate of Standards and Curriculum.

Ownership and sustainability of the CA programme

In order to guarantee the success of the programme it was important that the officers that were going to implement the programme had a sense of ownership of the programme and that the instruments and materials to be used were user friendly and built on the existing activities, materials and plans. This would guarantee sustainability. In this regard, the officers embarked on:

Preparation of assessment and monitoring guidelines for CA

This activity was done by a combined team of officers from Examinations Council of Zambia, Curriculum Development Centre, Teacher Education Directorate, Directorate of Standards and Curriculum, University of Zambia, Colleges of Education and District Inset Providers. The objective of the workshop was to prepare draft Assessment and Monitoring Guidelines for the CA programme

Preparation of assessment schemes for CA

A team of officers drawn from Examinations Council of Zambia, Curriculum Development Centre, Directorate of Standards and Curriculum and some Grade 4 and 5 teachers from each of the nine provinces were involved in this task. The officers were tasked to develop draft:

- Assessment Schemes for the CA programme. At Grades 5 and 6 in Mathematics, English, Creative and Technology Studies, Community Studies, Integrated Science and Social and Development Studies
- Assessment Specifications for End of Term and End of year assessment
- Sequential Assessment programme showing the different times when different aspects of the syllabus would be assessed
- Assessment formats for different learning outcomes
- Guidelines on the length and duration of assessment.

Preparation of assessment instruments, tasks, activities and marking schemes for CA

Officers drawn from Examinations Council of Zambia, Curriculum Development Centre, Directorate of Standards and Curriculum and Grade 5 teachers from each of the provinces and from the pilot schools were tasked to:

- Develop draft Assessment Instruments, Tasks, Activities and Marking Schemes in the six Learning Areas (Mathematics, English, Creative and Technology Studies, Community Studies, Integrated Science and Social and Development Studies) in line with the developed schemes and assessment specifications.
- Scrutinise and polish up the assessment tools, tasks activities and test items.
- Prepare assessment tools, tasks and activities to be used according to the developed assessment schemes.
- Develop Marking Schemes for each assessment tool, task and activity to be used by teachers at school level.

The Assessment and Monitoring Guidelines, Instruments, Tasks, Activities and Schemes were produced for discussion, amendment and enrichment by other stake holders.

Orientation of officers and head teachers from pilot districts and schools on the implementation of CA

Head teachers from each of the pilot schools, Provincial Education Standards Officers, Provincial Resource Centre Coordinators, District Resource Centre Coordinators and Statisticians from each of the pilot provinces and districts were targeted. These stakeholders were oriented on:

- Assessment and Monitoring Guidelines for implementing the CA programme
- Assessment Schemes to be used in the CA programme
- Different forms of Assessment Tasks, Guidelines and Activities to be used in the CA programme
- Assessment records and reporting procedures to be used in the CA programme
- Administration and management of the CA programme

Training of Grade 5 teachers from pilot schools in CA procedures

Training of Grade 5 teachers, School Inset Providers, Zonal Inset Providers and Guidance teachers in each of the pilot districts was done by officers from Examinations Council of Zambia, Curriculum Development Centre, Directorate of Standards and Curriculum.

- The objectives of the training were to:

- Orient the officers with the Assessment and Monitoring Guidelines for implementing the CA programme;
- Familiarize the officers with the Assessment Schemes to be used in the CA programme;
- Orient the officers with the different forms of Assessment Tasks, Guidelines and Activities to be used in the CA programme;
- Familiarize the officers with the assessment records and reporting procedures to be used in the CA programme;
- Orient the officers with the administration and management of the CA programme;
- Provide an opportunity for the officers to review the CA materials and give appropriate feed back about the suitability of the material to be used in the CA programme.

Development of parallel forms for Grade 5 pre-test and post-test instruments

The task was to:

- Develop the test specifications in the six learning areas based on the Grade 4 (pre-test) and Grade 5 (post-test) curriculum.
- Develop two versions of the pre-tests and post-tests based on the designed test specifications.

Materials production

Materials to be used in the conduct of CA have been produced by a combined team of officers from Examinations Council of Zambia, Curriculum Development Centre, Directorate of Standards and Curriculum. The materials were developed by teachers from pilot schools and some trained Grade 7 Item Writers. The materials were reviewed by Heads of schools, Standards Officers, District and Provincial Resource Centre Coordinators, Provincial and District Statisticians from the pilot provinces and districts.

The following have been produced and distributed to pilot schools, districts and provinces:

- Teacher's Guide
- Grade 5 CA Tasks in Six Learning Areas
- CA Class Mark Sheet
- CA Final Summary Mark Sheet
- Grade 5 Base Line Tests in the six learning areas
- Grade 5 Post Pilot Tests in the six learning areas.

Piloting

Piloting of the CA system started in January 2006. Baseline Tests in the six learning areas were administered in the second week of January to all grade five pupils in the 24 pilot schools and to all grade 5 pupils in 24 control schools. The pilot schools have received all the materials listed above while the control school only had the tests administered.

Grade 5 Post Pilot Tests in the six learning areas will be administered to the same pupils both in pilot and control schools in November 2006 in order to find out whether there will be any improvement in learning achievement of pupils in the pilot schools as a result of the use of CA as a learning and teaching strategy.

Challenges and constraints in the implementation of CA

The large class size in most primary schools is a major problem. It is common to find classes of 60 and above. Teachers are expected to mark and keep records of the progress of the learners. Work related to CA takes a lot of time for teachers and as a result teachers get concerned that time spent on remediation and enrichment is too much and as a result they would not finish the syllabus. There is a belief amongst teachers that CA involves too much work

CA will not be successfully implemented if there are inadequate teaching resources/ equipment in schools. Teachers need materials and equipment such as stationery, computers, photocopiers and electricity.

Use of the CA data

The data generated from continuous assessment will be useful in assisting teachers plan work for individual pupils. It will also assist teachers in identifying the unique understanding of each learner in a classroom. It will also inform the pupil of the level of instructional attainment. Continuous assessment will help target opportunities that promote learning and help to reduce learner anxiety associated with examinations.

Conclusion

The idea of CA is to have no child left behind and almost every child should shine. CA aims at involving Teachers in assessment. CA will help to reduce pressure on children preparing for examinations as the marks obtained from CA will count towards the final mark for selection and certification. Class sizes should be monitored and reduced to manageable levels. CA focuses on mastery learning and improves holist education of learners. It is therefore an issue of standards and the success of the programme depends on how Standards Officers apply themselves to its successful implementation.

Training in CA should be continuous and incorporated in pre-service training of teachers.

The fallacy of measuring mathematics 'competencies' in a context of a 'high stakes' external examination

Thabiso Nyabanyaba, University of the Witwatersrand, Education Policy Unit

Abstract

Recent curriculum reforms have led to a wider variety of assessment practices, including more novel methods of assessing 'higher-order' skills and more examination items that are based on 'realistic' contexts. My study of Basotho students' epistemological access to school mathematics revealed that, in the context of 'high stakes' examinations, students from different sociological background are positioned differently as they attempt these examination items with novel requirements and standards. In this paper, I will use extracts from in-depth interviews with three experienced Basotho[1] teachers to highlight the failure of these well-intended reform initiatives to attract the critical engagement of teachers because of the teachers' peripheral involvement in such initiatives. This raises serious concerns around the legacy of post-colonial education in which school-leaving examinations are still set outside of some African countries. I will also use extracts from interviews with Basotho students selected from a variety of socio-cultural backgrounds to draw attention to the tendency for high stake external examinations to promote socio-economic access rather than epistemological access. The findings of this study raise questions about reform initiatives driven by international examination boards and whether the standards derived from such initiatives can accurately measure individual learner 'competencies' in local contexts.

Introduction

There is growing concern that schools are not producing citizens that are able to apply their knowledge across a variety of contexts. The argument that in this *information age* the

[1] Refers to the people of Lesotho.

knowledge, skills and behaviours needed for students to achieve at high levels require the ability to access, interpret, analyse and use information for making decisions is very persuasive (Bond, 1994). It is hard to refute the claim that today's world requires that citizens and the workforce should possess more than basic reading and arithmetic skills. Advocates of the standards-based reform movement argue that dissatisfaction of employers and leaders in higher education about the quality of knowledge and skills presented by graduates of public schools prompted calls for higher standards in public education (Reville, 2002). Many believe that the impetus for the current wave of standard based reform can be traced back to the alarming United States federal report entitled *A Nation at Risk* by the National Commission on Excellence in Education in April 1983. This report claimed that American students were not studying the right subjects, were not working hard enough, and were not learning enough and their schools suffered from slack and uneven standards.

Internationally, assessment of learner competencies has played a significant role in the hegemony for standards-based reform initiatives. In a number of countries such as Australia and the United Kingdom the growing concern with the monitoring of standards are accompanied by a great deal of focus on student outcomes (Rowe, 2000). In the latter country, the impact of standards-based policy focus has resulted in an obsession with comparing schools' average achievement scores on tests and public examinations. Publication of these 'league tables' of examination and national curriculum tests results is more explicitly said to be intended to help inform parental choice about schools, and less explicitly about public demand for accountability and the maintenance of educational standards (Hill, 1994). Assessment has been referred to as 'arguably the most powerful policy tool in education (Broadfoot, 1996: 21). However, in developing countries, examinations are often the sole means of judging educational standards and are rarely ever used as a means for improving pedagogy (Kellaghan & Greaney, 1992; Pennycuick, 1993).

Arguments against the surge of reform in school mathematics assessment practices include caution against initiatives based on political-economic ideals rather than research. Criticism of standards-based reform and the surge of outcomes-based education has also focused on the influence of market-oriented ideals for competitiveness which often background education research (Rowie, 2000; Jansen, 2002). Another element of recent reform initiatives that has been under scrutiny lately has been the increasing inclusion of 'realistic' contexts in school examinations. Findings from this study and other studies (Cooper & Dunne, 2000) show that the inclusion of 'realistic' items can lead to an underestimation of students' mathematical competencies. However, what emerges here is the profound implication that the 'high stakes' attached to examinations for selection in post-colonial countries pervades any epistemic explanation for access to school mathematics through 'realistic' tasks.

Background

To fully understand the 'contextual parameters' of assessment practices in Lesotho, one needs to have some insights into the colonial legacy. A former British Colony, Lesotho (previously Basotholand) is completely surrounded by its more economically viable neighbour, South Africa. The secondary school leaving examinations in Lesotho were run by the Joint Matriculation Board of the Universities of South Africa (JMB) until 1961. Since then, the University of Cambridge Local Examining Syndicate (UCLES), based in the United Kingdom (UK), has been offering the Cambridge Overseas School Certificate (COSC)

Examination as a school leaving examination for Basotho students (Pule, 1991). Although the marking of this examination is now being done locally in Lesotho, the actual examination itself is still set in the UK. All examinations in Lesotho, from the Primary School Leaving Examination to the COSC, are external in that no internal assessment contributes to the student's final mark. The high stakes attached to such an examination for selection, and not for feedback, and the sociological inequities being perpetuated through such a system, shall become a major focus of this paper.

School mathematics is a particularly rich research subject because the subject is often used to distribute social opportunities for young citizens. School mathematics has always been a recontextualisation from the discipline of mathematics and that of education, and what counts as school mathematics is as much a matter of the discipline of mathematics as it is a question of different philosophical and political orientations. The various philosophical movements such as the ethnomathematics movements; the Realistic Mathematics Education movements; and the various political imperatives such as those describes by Jansen (2002) further recontextualise what counts as school mathematics differently in their regards for what should constitute the school mathematics practice. Whatever their philosophical basis, their political imperatives, or their educational intentions, Cooper and Dunne (1998) argue that reform initiatives such as the inclusion of realistic contexts in the school mathematics curriculum in the UK has followed on official documents such as the Cockcroft Report (Cockcroft, 1982), which gave sound epistemological arguments for the inclusion of 'relevant' contexts rather than findings from empirical research. In this argument they are supported by Gipps (1996) who argues that politicians have the inclusion of such contexts as a pre-determined path. A particular concern here is with regard recent curriculum reforms that have led to a wider variety of assessment practices, including more novel methods of assessing 'higher-order' skills and more examination items that are based on 'realistic' contexts.

Methodological issues

The focus of the paper shall be on Basotho teachers' and students' positioning in relation to reforms reflected in the O level mathematics examination as revealed by the interviews. Two education officers who were most intricately involved in secondary mathematics were interviewed regarding the extent of support given to teachers. However, the focus of this study was on teachers' professional identity and positioning as indicated by the extent to which they critically engaged with such reforms. This analysis was undertaken through an interview with three purposively selected Basotho teachers, whose wide experience would have been expected to assist them relate to and engage with reforms. The second level of analysis of the impact of these reforms was on how Basotho students are positioned in relation to the trends and what they attribute their positioning to. This was undertaken through interviews following an administration of a recent examination item in three groups of students from three schools that typically draw from students from different socio-economic backgrounds.

The analysis began with a study of the changes in the COSC mathematics examination and that analysis led to a development of a language of description that illuminated the impact on students' access and performance. The analysis then focused on Basotho teachers' and students' positioning with regard the reform initiatives as revealed in the interviews. In

conclusion, the paper argues that the post-colonial context of external examination continues to pervade 'noble' reform initiatives aimed at higher-order thinking skills and the ability to apply knowledge across contexts.

A language of analysis

An investigation into the embedding of examination items in real-life contexts introduced a very interesting dimension to the issue of standards. However, it became important to develop a language for analysing the requirements of this novel demand on Basotho students. Empirical work done by Dowling (1998) developed a sophisticated language that allowed for an analysis of the inclusion of 'realistic' items into school mathematics textbooks and the negative impact of that inclusion on the very activities they were trying to affirm. The term 'esoteric' has been used by Dowling (1998) and Cooper and Dunne's work (1998; 2000) to distinguish between mathematical and non-mathematical activities. In their work, Cooper and Dunne (2000) define a 'realistic' examinations item as one that contains either persons or non-mathematical objects from 'everyday' settings. Otherwise the item is referred to as 'esoteric' (or in the grammar of mathematics). Cooper and Dunne (2000) use a sociological approach to develop a language of description around the analysis of the difficulties that children experience around mathematical items embedded in 'realistic' contexts. They draw on Bernstein's (1996) 'recognition rules' and 'realisation rules' to describe differential access across social class. Cooper and Dunne (1998) define the 'recognition rules' as the means an acquirer employs to "recognise the speciality of the context" they are in and the 'realisation rules' as the means of allowing for the production of "legitimate text". This ability to recognise the speciality of the context can also be referred to as the possession of 'educational ground rules', a theoretical construct originally used by Edwards and Mercer (1987) but also used by Cooper and Dunne (1998).

Cooper and Dunne (2000) illustrate how, given a 'realistic' item as opposed to an 'esoteric' item, children from working class families remain tied to their 'everyday' meanings in a manner that is self-defeating. In one illustration, a student, 'Mike,' is required to sort rubbish "from the sports field into one of two circles drawn". The 'rubbish' to be sorted included a newspaper, a can of drink, a bottle of mustard, a pen and a carton of milk. The marking scheme indicated that it was acceptable for children to sort "by shape of container, by being edible or drinkable and so on" (Cooper and Dunne, 2000: 50). Mike's sorting is very much related to the materials themselves as either paper or metal (he wrote "meatle"). Another student, 'Diane,' responded by illustrating the recognition of both material and metacognitive sorting. She also demonstrated an awareness of the priority attached to the metacognitive level in school. Diane's sorting prioritised 2-dimensional over 3-dimensional diagrams, although she also illustrated an awareness that sorting can be done at a material level. Further investigation revealed that Mike could also make appropriate recognition of appropriate practice but would not illustrate the explicit priority that could Diane. As a result, such items lead to an underestimation of the working class children's mathematical competencies.

The distinction that Cooper and Dunne (2000) make between 'realistic' and 'esoteric' question provided me with a useful framework for analysing general patterns in the O level examination over the past 15 years. For example, the difference between these two opening

O level mathematics questions, one set in 1984, and the other set in 1996, illustrates the move away from esoteric to realistic contexts.

1 *(a)* Find the value of $(1\frac{1}{2} + \frac{2}{3}) \times 1\frac{1}{5}$, giving your answer as a fraction in its lowest terms. [2]

(b) Evaluate 5.4 + 4 x 0.3. [2]

(c) Calculate 6% of \$5450. [2]

(d) Express 42 cm as a percentage of 1.05 m. [2]

(4004, 1984)

1 **(a)** 100 g of beans contains 4.7 g of protein.

Calculate the mass of protein in 3.3 kg of beans. [2]

(b) To make a chicken and mushroom pie Mrs Lee requires 550 g of chicken and 125 g of mushrooms.

Express the mass of the mushrooms as a percentage of the mass of the chicken, giving your answer correct to the nearest whole number. [3]

The author of a cookery book uses the formula

$$T_2 = T_1 \sqrt[3]{m^2}$$

to make up a table of cooking times for meat.

In the formula

T_1 is the time in minutes to cook a mass of 1 kilogram of meat.

T_2 is the time in minutes to cook a mass of m kilograms of meat.

Calculate T_2 when $T_1 = 20$ and $m = 2.5$. [3]

(4004, 1996)

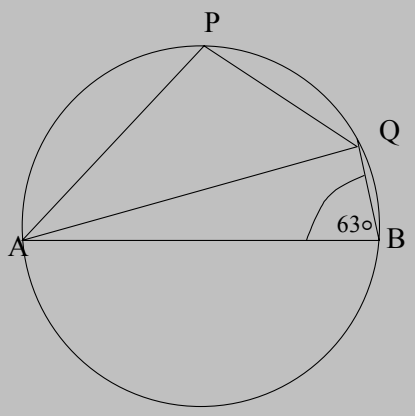
In a context of examination, however, the distinction between 'realistic' and 'esoteric' tasks needed to be extended further to capture the more textured differences, both in terms of the contexts and the demands, displayed by the questions. The growing concern for higher-order skills is also illustrated by these two sets of questions on angle properties of a circle:

2 *(c)* P and Q are points on a circle whose diameter is AB.

Given that angle ABQ = 63º, calculate

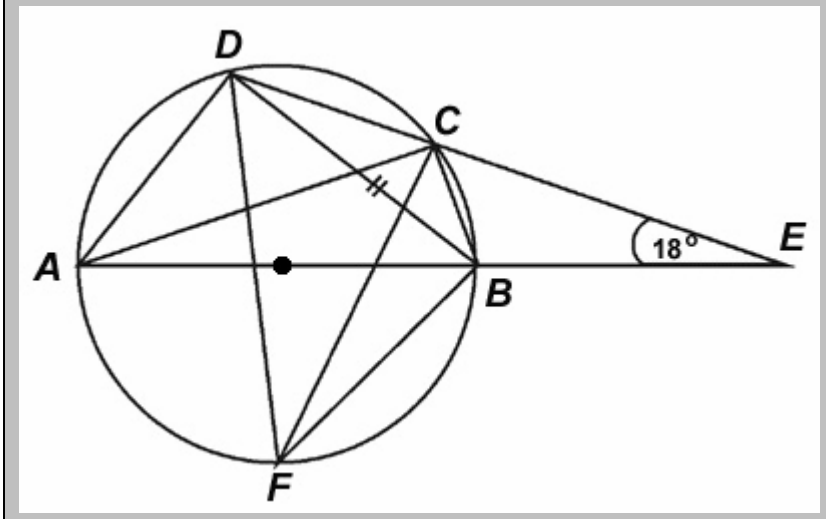
(i) $A\hat{P}Q$,

(ii) $Q\hat{A}B$. [2]



(4004, 1984)

4



AB is a diameter of the circle AFBCD shown in the diagram.

E is the point on the AB produced, where BD = BE and $A\hat{B}D$ = 18º.

(a) explain why C$\hat{F}$B = 18°. [2]

(b) Find A$\hat{B}$C. [2]

(c) Show that BD bisects A$\hat{B}$C. [2]

(d) Given also that B$\hat{D}$C = 51°, calculate F$\hat{B}$C. [1]

Concepts that I found further useful in analysing the questions in the examination were those applied by Blum and Niss (1991) in their argument for 'proper' applications in school mathematics. In differentiating between an 'exercise' and a 'problem', they define a problem as:

> [A] situation which carries with it certain open questions that challenge somebody intellectually who is not in immediate possession of direct methods/procedures algorithms etc. sufficient to answer the questions.
> (Blum and Niss, 1991: 37)

Otherwise, if one is simply rehearsing the application of some learnt algorithm, then the situation becomes an exercise. The authors, of course, concede that a problem is relative to the person involved: what to one person is a problem, to another is an exercise. The following question which appeared in the 1998 O level mathematics paper begins to illuminate a question that could for many students be novel both in the context and demand.

7 [The value of π is 3.142 correct to three decimal places.]

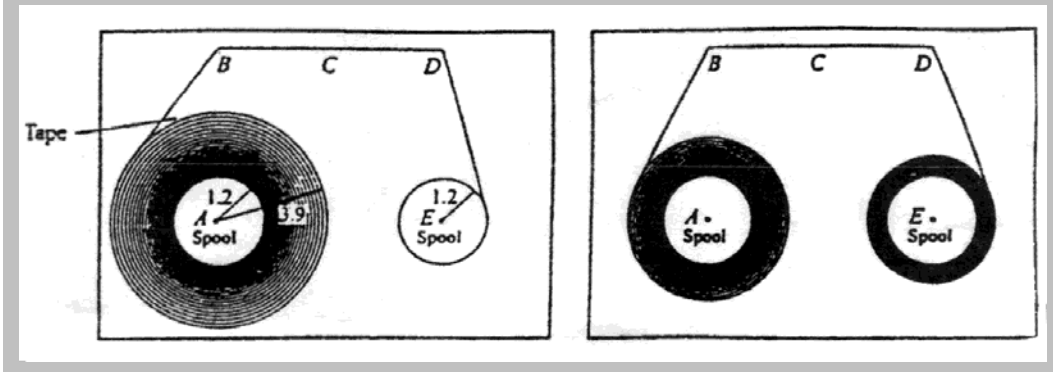


Diagram I **Diagram II**

Diagrams I and II represent the cross-section of a video cassette.

A tape runs from one circular spool, centre A, past B, C and D, to a second circular spool, centre E. Each end of the tape is fixed to one of the spools, both of which have a radius of 1.2 cm.

Initially as much tape as possible is wound on to the spool with centre A.

It is represented on Diagram I by the shaded area, whose outer radius is 3.9.

(a) Show that this shaded area is approximately 43.3 cm^2. [3]

(b) Diagram II represents the situation when some of the tape has been wound onto the second spool. The total shaded area remains unaltered, so that it is always approximately 43.3 cm^2.

At a certain time there are equal lengths of tape on each of the spool.

Calculate the outer radius of the tape on one of the spools at that time. [4]

(c) The tape runs at a speed of 23.4 millimetres per second past C.

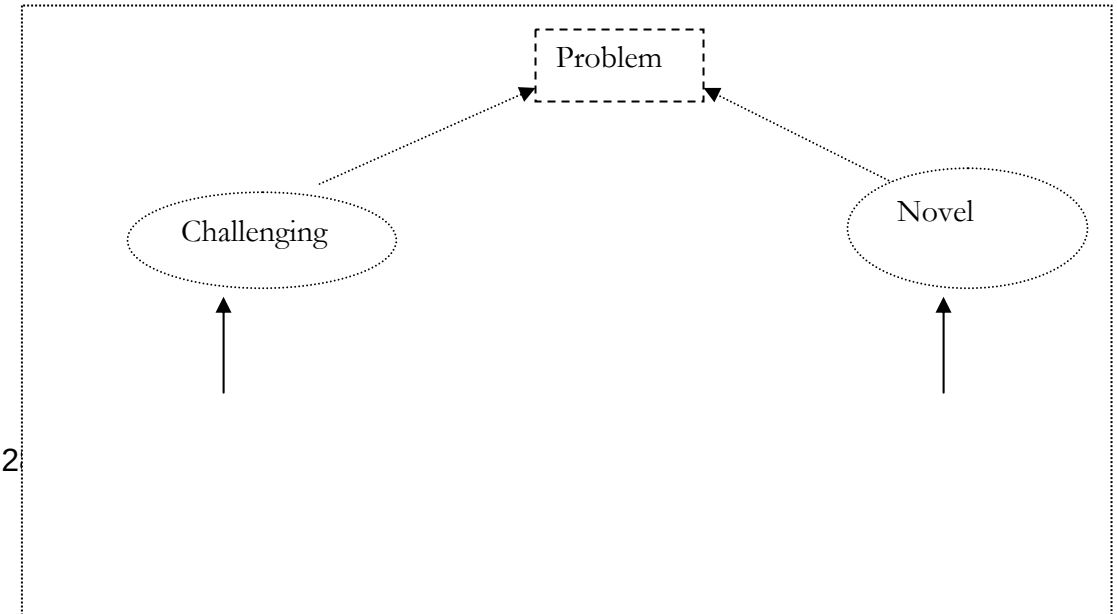
It takes 3 hours for the tape to run from one spool to the other.

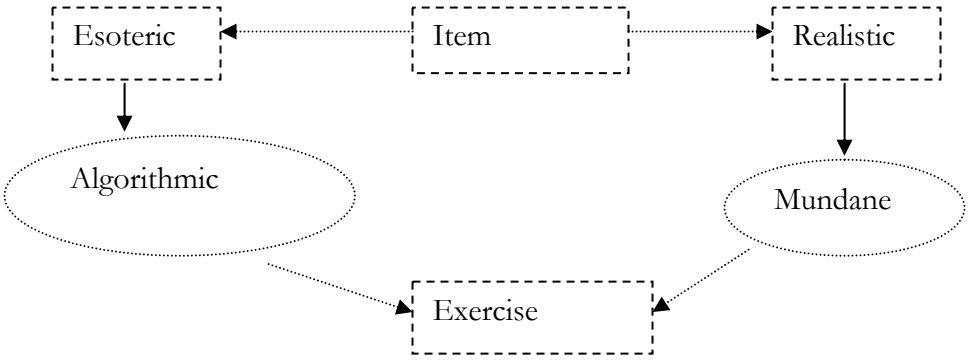
Show that the length of the tape is approximately 250 metres. [3]

(d) Using the results of parts (a) and (c), calculate, in millimetres, the thickness of tape. [2]

This analysis begins to challenge the conception that 'standards' are merely high or low. It strongly suggests that it is not merely a vertical order; it also brings in the intersection between the novelty of the demand made in the question and the possible unfamiliarity of the context in which the question is embedded. The diagram below describes what emerged for me as a language of distinguishing the various impacts of the curricular trends.

Emerging language





In short, a question can either be 'realistic' or 'esoteric' depending on the context. Depending on their mathematical demands some esoteric questions are able to differentiate between different mathematical competencies to a greater or lesser extent. On the one hand, they differentiate to a greater extent if they are so challenging especially in their mathematical demands that they become a problem. On the other hand, they differentiate to a lesser extent when they have been so well rehearsed by students that they only demand mere algothrims on the part of the students. As noted before, whether a question is challenging or algorithmic is not solely dependent on the question, but is also a function of the learner's relationship with the question.

A realistic question can be so mundane that learners can recognise the mathematical demands without any reference to the 'realistic' context. Some questions are mundane in and by themselves. Again some learners may be so familiar with the context of the question that the context is rendered mundane. However, the contexts of some 'realistic' questions may be so novel that they require close reference to the context and may even become inaccessible to those students who are unfamiliar with their contexts. Note that an additional problem with a 'realistic' question is that whether it is 'mundane' or 'novel', it can still be challenging or algorithmic with its mathematical demands on the learner. But much of what determines the degree to which questions are accessible to learners, whether in their contexts or in their mathematical demands, depends on the internalised 'ground rules' of the learners. These internalised 'ground rules' are evidenced through the rounder parts of the diagram. An item can be either esoteric or realistic and can even be a problem or an exercise on and by itself. However, largely depending on the 'ground rules' internalised by students the item can be challenging or algorithmic and its context can be ritualistic or novel. In a more developed language employed by Cooper and Dunne (2000) these 'ground rules' are generally the recognition and realisation rules. Therefore, in the context of 'realistic' items, particularly those that are novel, the standards concern is not merely whether they are highly demanding, but whether they are also familiar. As shall be noted later, how different learners deal with these novel 'realistic' question differs so significantly as to introduce a greater element of socio-economic background than such reform initiatives would acknowledge.

Investigating teachers' peripheral positioning in a context of external examinations

I focus in this section on the extent to which teachers in Lesotho have developed a strong sense of professional identity, particularly in relation to recent curricular reforms. There was disagreement between the accounts of the education officers and the teachers interviewed regarding the extent to which teachers received support on curricular reforms. Education officers described workshops in which teachers were 'alerted' to changes in the examination format but conceded that there was neither document nor dialogue regarding these reforms. Teachers argued that even the workshops were either too few or inadequate for a proper exploration of these reforms. However, what was of interest to me was the extent to which teachers were able to engage critically with some of the curricular issues that I put before them in probing for their reflections, for instance on the trends in the mathematics examinations.

Even in the most participatory education system, some teachers might find it difficult to articulate curricular issues. Moreover, the examination-focussed nature of the curriculum in Lesotho meant that many teachers were more concerned with preparing their students for the examination than reflecting upon its appropriateness. Also the 'inadequacy' of in-service programmes for teachers could mean that many do not have ample opportunities to talk about curricular issues. Nevertheless, one would expect these particular teachers to be able to talk about the two examinations and be much more involved in issues arising, not least because all of them had a great deal of experience, two of them as long-time markers in O level mathematics.

Regarding recent trends School C teacher vaguely referred to how there was a gap between local demands in the classroom and the requirement to apply their skills in the O level examination, but was unable to describe the trends in terms of curricular reforms:

> **CI** What do you think might be the cause [of the gap]?
>
> **CT** I think … I don't know. There's more … I don't know. They do have the basics. The syllabus expects them to have the basics. And I think they basically study, and we teachers actually train them for the examination. That is why they get into COSC not being able to apply those basics because COSC mathematics is applications.

Another teacher was only able to imply that such reforms led to some difficulties regarding accessing the demands, particularly for second language speakers of English:

> BI What changes are you aware of in the mathematics syllabus, reflected through the COSC examinations?
>
> BT … what I've also noticed is that the questions are long and verbose; two, they're too diagrammatic, OK? Some of the diagrams, you know, you need actually to sit down take some time before, as a students, before you can easily identify what the question is about.

While there was a general feeling that such trends make it difficult for Basotho students to access such question, all three teachers, irrespective of the school in which they worked, expressed helplessness regarding what was happening in the examination. School B teacher continued to describe how he at first tried to write to the locally set Junior Certificate (JC) examinations board, only to be snubbed:

> BI … have you ever been compelled to write remarks about the COSC paper?
>
> BT No!
>
> BI Why not?
>
> BT Umm.. one, it's not local. Two, I thought I should start from home. That was why I wrote for this JC. There was no reaction. In fact, they didn't write back at least saying we have received your something, and therefore thank you, or we shall look into that, or some … at least, a word of encouragement, yah, never came. And I felt it was not worthwhile if they don't recognise your efforts. They don't appreciate what you're doing, yah. So I didn't write anything about COSC.

The other two teachers admitted that they did very little - or nothing at all - to comment on or contribute to even the local Junior Certificate examination trends. For example, the teacher in School C admitted that she had come across questions that she was unhappy with, yet she did nothing. She argued that all she felt she could do was prepare her next students better:

> **CT** If I see … if there's a question I didn't expect or surprises me, then the following year I try to look into it.

She felt that as a teacher she has no control over what happened in the examination, except to make her next students aware of the possibility of such a question. Even the teacher in School A, who had been involved in the marking of the O level mathematics examination for a long time, expressed the feeling that she would not critically engage with the setting of questions in the O level examination, at least not as a 'mere' teacher:

> AI … Do you ever complain; do you ever communicate with the setting people?
>
> AT That we do when we are writing the report after marking. That is the only time that we usually communicate with them.
>
> AI But as an individual teacher you never do it?
>
> AT No, I've never really done it!
>
> AI Why not? Why didn't you feel you have the right to write from your position as a teacher?
>
> CT As a teacher … I have no reason.

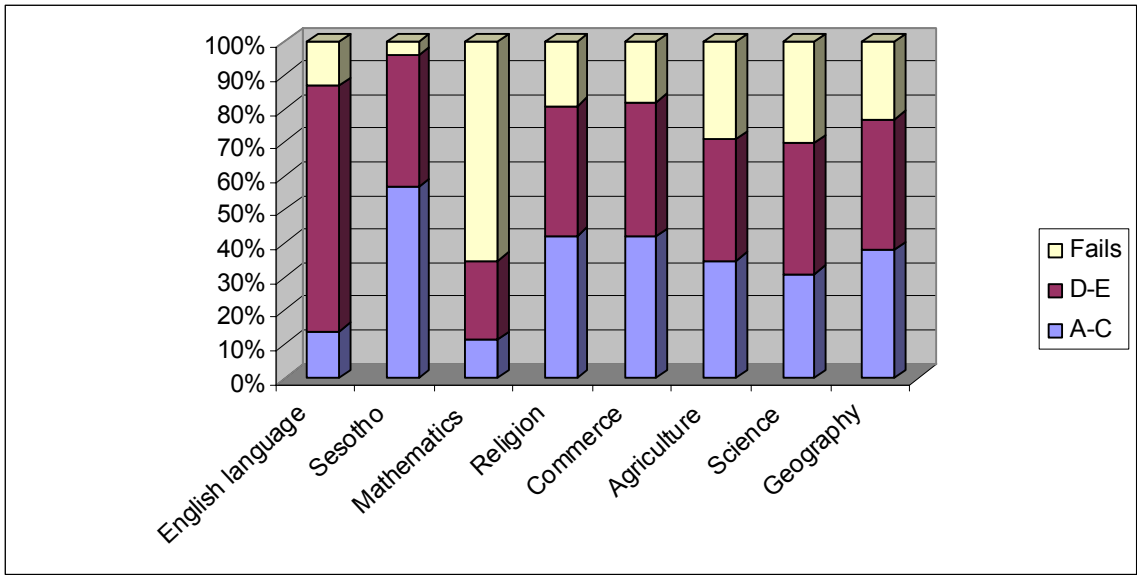
It is apparent that teachers' professional capacity has not been legitimated by the emphasis on summative and external examinations in Lesotho. Even in the local examination the system is relatively detached from teachers, acting as an independent agency. Remarks such as the one made by Teacher BT—that it did not make any difference to write to the local examination board about teachers' unhappiness—indicates the peripheral position of teachers relative to curricular developments in Lesotho. Descriptions of how secondary mathematics teaching is limited to 'exposition-example-exercise' abound in Lesotho's education literature (Kokome, 1990, Moru, 1994). However, this study indicates that Lesotho's examination driven curriculum might account for some of the absence of reflections and innovative teaching approaches by teachers. The disjuncture between an externally-based assessment system in Lesotho and teachers' practice does not help improve engagement of pedagogic practice. That even the local examination does not engage with teachers' comments and there is a general feeling that it is fruitless to communicate with the examination authorities does contribute to the peripheral positioning of the teachers. This absence of a teachers' voice must affect the teachers' legitimacy as players in the system.

This study suggests that teachers are almost incapable of reflecting on curricular issues. The system inherited by Lesotho, where events in the curriculum, which is supposed to be delivered by local teachers, do not concern or involve them, only results in their putting in very little reflection into their practice. Because of the post-colonial context and teachers' peripheral position, teachers tended not to be able to think about or engage with curricular trends. As a result, there was an *absence* in the teachers' talk of indicators of their professional identity, not only with regard to conditions of service, but also to curriculum issues.

If, as this section illustrates, teachers feel themselves at the periphery of curriculum and professional developments, it would be surprising to find any more 'presences' with the students. As the next section illustrates, there were just as many 'absences' in the students' positioning relative to 'realistic' items. These occurred irrespective of the quality of student-intake the school attracted.

Investigating positioning in a context of 'absences'

The performance of Basotho students has generally been described as poor, particularly in mathematics (ECOL, 1999). Failure rates in O level mathematics in Lesotho currently stand at 65%, as illustrated by the chart below.



In contrast to these high failure rates in mathematics, failure rates in all other subjects commonly taken by Basotho students for their school leaving examination, stand at less than 30%. Even in the foreign language (English language), the only notable concern would be for the modesty of the performance, with most (72%) learners scoring a modest pass (D-E) and only 14% attaining credits (D-E). However, it is fast becoming the story even in developed countries that performance does not tell all, especially about who is 'being left behind' (Darling-Hammond, 1994). The performance of Basotho students certainly does not tell the story of what is going on in schools and why they are happening. Using the language I developed earlier it may not be clear why a question such as question 7 is a problem until one has had insights from students. For example, a question can be a problem merely by being so novel or just by being so challenging in its demand that it becomes inaccessible to the majority of students. Some questions may become problems because they are both novel in their contexts and challenging in their demands.

In order to further investigate what the problem with such questions was and why not everyone failed the examination, I assigned question 7 on a cassette (cited earlier) to several groups of students who were in a context where I could follow up on their positioning. This task was assigned to completing students in 3 schools in Maseru, the capital city of Lesotho. The selection of the schools was opportunistic in that I knew the teachers and expected that they would be willing to provide me with insights into their positioning and allow me access to their students. But I was also conscious in choosing the three schools that they, on the basis of the 1998 and previous results, covered a spectrum of performance to below average, average and above average results. The task or scenario was administered to students in the second half of 1999 and was then followed by interviews with students. Both the teachers and students testified that the task had never been administered in class. Even those students who had access to past examination papers were likely to have ignored the task as the interviews cited here will reveal.

Within the broader study, the task above was analysed more extensively in terms of both the authenticity of the 'realistic' context and the depth of the mathematical demands

(Nyabanyaba, 2002). In terms of its authenticity the context raises questions about how familiar some Basotho students would be with a context of a video cassette. Those who are unfamiliar with the context of a cassette, perhaps from the rural area, would be at a disadvantage regarding how a cassette operates. This would have serious consequences regarding their interpretation of the relationship between Diagram I and Diagram II which would be a given for anyone who is familiar with the context. The issue of the relationship between the diagrams was raised and will be further discussed in relation to students' recognition of the mathematical demands of the question later.

Although, the interviews were conducted in English, students were allowed to respond in mother tongue (Sesotho). Only the English versions are provided here but the original transcripts, including where students responded in Sesotho, are available in the thesis document (Nyabanyaba, 2002). In the transcripts provided below, **I**, represents the interviewer who was also the researcher. Respondents are distinguished according to whether they belong to school **A**, **B** or **C**, whether they are students (**S**) or teachers (**T**). Where necessary, the respondents are further numbered. For example as **BS1** would be the first respondent within this extract which was from an interview of students from School B.

Only about 18% of the students had returned their scripts with more than one sub-question (e.g. (a)) out of four attempted. This situation constitutes a high level of absences which illustrate just how reluctant students were to complete the task, unless explicitly guided as was the case in School C. In this extract, I had asked students why so many of them had given very limited to no responses to the task.

> **BS** No, we didn't even try it because when you see that something is very difficult, you don't know how to start.

But the nature of that 'difficulty' became subtler as I probed further for what constituted the difficulty. The next section looks into what led to these 'absences' and reluctances as revealed by more interviews with the students.

Basotho students: recognition and realisation

I extended my interviews with students so as to probe for reasons why they had not answered this question as well as what specific 'difficulties' they experienced. Important issues were raised about the deep 'realistic' context of this question, and especially how it affected Basotho students. The following extracts are from class or group interviews:

> AS The diagram itself is confusing.
>
> AI What is confusing about it?
>
> AS What is happening there.

'What is happening there' and what students have to figure out is as much the mathematics concepts embedded as the relationship between the two diagrams. In school B, the students stated more explicitly that the tape context was very confusing for them, particularly how the different diagrams related to the context.

BS ... and this way it confused me when it says they what ... the total shaded area remains unaltered, and yet the tape moved to the other side.

The context of the question made very complex demands on Basotho students in terms of interpreting the relationship between the diagrams from their experiences. The context also made very specific linguistic demands for the students to hold several pieces of information together in a second language. In fact, the students indicated that they found the whole context quite intimidating. In School A, I literally had to go over the whole context before I could even discuss the question or its demands. However, what is worrying is that once the context had been discussed and the demands demystified many students were able to demonstrate the mathematical demands required by the question. This was interesting because it could point towards the difficulty students were having recognising those mathematical demands. Such lack of recognition, which can impede the realisation of mathematical competencies, can also mislead examiners into assuming that they point towards a lack of competence. Therefore, while the performance of students can be described at a more superficial level as poor, there do appear to be more subtle reasons for any description of these 'poor' performances. This demonstration, and the fact that students in School A generally continue to do well in spite of their inability to recognise the mathematical demands, was an interesting difference to investigate. Just how the lack of recognition might be disadvantaging some students more than others is the focus in the next section. In Cooper and Dunne's (2000) study, the recognition rules were easier to relate to the social class because there were 'presences'. In the 'absences' of the 'presences', I then looked more closely at what were the different dispositions of students in relation to the 'realistic' questions and how the dispositions or 'habitus' impacted on different students' performance.

Students' habitus: differential socio-economic status

There were various indications that students' socio-economic backgrounds were vastly different. But one very interesting element of the backgrounds of students was the emergence of the privatisation of education in Lesotho. This privatisation was evidenced by some students attending extra-lessons after school. These extra-lessons or evening classes are often referred to in Lesotho as 'tutorials'. The tutorials or 'evening classes,' which focus on examination preparation, provided a strong advantage to those students whose parents could afford to send their children to them. In School A particularly, the teacher described the practice amongst students of attending evening classes, as well as that of purchasing past examinations papers, as prevalent. In our discussion of the 'realistic' item I had given to her students, the teacher revealed these practices:

AT They have all the question papers and they discuss. But rarely … they rarely ask us [the school teachers] really, you know the outside question, other than the questions that we do with them or we ask them in class.2 Very few of them would come. Sometimes they do them in their evening classes.

[2] The teacher here means that there are a variety of questions that students have purchased from the Examinations council of Lesotho (ECOL). Such questions, which are 'outside' the activities of the class, are discussed in the tutorials.

AI What would you say would be percentage of students attending evening classes?

AT Here?

AI Yah.

AT Um-m, a big percentage!

Although less prevalent, the practice of purchasing past examination papers existed in the other two schools. However, in the other two schools which attracted students of parents with poorer socio-economic backgrounds, students would not be able to afford to purchase the papers. Even with attending 'tutorials', there were fewer to no students in those two schools.

CT There're not that many! But there're students who do attend private tutorials.

In school B the students' inability to attend the evening classes also had to do with the fact that they attended school until late.

BT Eh, unfortunately we don't …our students don't do it.

It appears from this analysis that not as many in School B and C would have afforded the tutorials as in School A. In fact, students in School B in particular requested me to supply them with past examination papers, unlike those in School A who, as the teacher maintained, had purchased their own. The value attached to these tutorial and the purchasing of past examination paper by students was an interesting finding because it illustrated differential resources beyond that which schools were able to distribute. Students who could afford to could then purchase past examination papers and attend evening classes. It is clear that these distributed advantages were very closely connected to families' economic and educational backgrounds (i.e. favouring School A over Schools B and C). This for me formed a very distinctive cultural capital, which was distributed differentially. These advantaged students were then able to make informed choices about what they should do in examination situations. Asked why they were reluctant to attempt the 'realistic' item, the students indicated that it would not normally be their choice of question in an examination:

AS1 I never even bother with such questions because I think there are enough questions in Section B … at least, there must be some graphs questions which will appear. And I have tried to make sure that with graphs, I attempt everything. It's because here in school we haven't done this topic; even there [in the tutorials], we haven't done this topic, those which refer to … Normally when I open a question paper, any question that is based on this topic I never … I just look at it and then proceed to graphs and perhaps vectors.

I OK? [Referring to the other student]

AS2 Yes, I never do this one.

I You don't do it at all?

AS2 Even when I revise, … even on my own, I never do it at all. That's because even when I attempt it, I fail.

The students have received extensive support on their preparations for the O level mathematics examination. These students who have attended tutorials would have been exposed not only to the mathematics, but also the specific context of examination. This awareness translates into a strong cultural capital that guides the students in the specific context of such high stakes examinations as the O level mathematics examination. This awareness emerged only in School A, where the students were able to articulate their awareness of the choices they had. The students in Schools B and C could not even articulate their position in relation to 'realistic' questions. In general, because the students come across the 'realistic' trends in a context of external examinations, the epistemological and motivational intentions are backgrounded by the high stakes associated with external examinations.

The examination is external and as removed from the teachers as it is from the students. In the context where many students find mathematics difficult, there are those who bring a strong socio-economic background into the examination. This background provides a strong cultural capital borne out of extensive preparations that the students' parents have been able to put them through. Such students show a great awareness of the choices they have in the examination and this clearly puts them at an advantage over the difficulties that all others experience with such novel demands and 'realistic' contexts as reflected examination question 7 on a cassette. These differences among students, generated by their generally poor backgrounds, some strong socio-economic backgrounds, and a few strong individual drives, is what I will refer to as the 'habitus' developed by the students. Within the 'habitus', the socio-economic backgrounds appear to play a very strong role in distributing advantages and disadvantages among students.

Conclusion

The concern for educational standards is very interesting in the contexts of high stakes examinations systems in developing countries. Particularly, in post-colonial countries such as Lesotho where school leaving examinations are still set outside the country, teachers' professional identity and practices are marginalised by the extent to which these curricular reforms are removed from them. While the marking and grading of scripts locally are events that are beginning to engage the professional judgement of teachers, the high stakes associated with these examinations means that teachers have as their prime concern the preparation of students for the crucial examinations that serve to determine their entire futures. Being entirely summative and external, the examination is as removed from the teachers as it is from the students. More meaningful and relevant mathematics practice is not necessarily contradictory to good examination preparation. However, the authority of teachers having been removed from them by an entirely summative examination, meant only for selection, they can only act as conveyors of knowledge legitimated by the examinations-driven education system.

An analysis of the curricular reforms reflected in Lesotho O level mathematics examination indicates that the standard embedded in the questions in only about high or low demands. It is quite intricately related to the socio-economic background and the cultural capital they

bring to bear on their positioning. What is much more significant about what the post-colonial context of this study adds to the concern about standards in school mathematics is that the context of high stake examinations pervades any noble curricular intentions such as the inclusions of 'realistic' items. It is all about passing the examination and students with a stronger habitus have the ground rules to select out questions with more confidence. These ground rules are not found in schools but are inherent of learners whose parents can take them through the 'private' tutorials that are becoming a central part of education in developing countries. The advantages come not from an 'epistemological' concern to show what one understands, but an almost conscious admission that the context of Lesotho is one in which everyone is doing poorly. In such a context, the recognition and realisation rules so unambiguously displayed by Cooper and Dunne's (2000) students are only partly about the awareness of the boundaries between the 'realistic' and 'esoteric' contexts. The recognition and realisation rules are about an awareness of the contexts they are in, the context of high stakes examination and the use of the disposition or habitus to select out confidently so that one maximises one's performance in an examination for selection. It is indeed not about the curricular reforms but about the high stakes associated with examination for selection. Whatever reform mathematicians will come up with will always be backgrounded by this context of high stake examinations.

As with most sociological problems the implications for practice are complex. Part of the problem might be addressed by examiners taking more care to evaluate the questions they set so that the do not obfuscate the mathematical demands. Teachers have to be brought into a more critical role regarding their positioning in relation to the curricular trends. More significantly, the impact of the external summative examination is even more significant in its impact on the distribution of social opportunities for young people in Lesotho. The obvious main solution would appear to be through an assessment practice that is geared not only towards selection, but also meant to provide feedback to the teachers through a valuing of formative assessment practices. Unfortunately, formative assessment in a context of developing countries is not a simple solution. It requires a great deal of human and material resources and therefore such a process needs to be carefully monitored so that it does not bring in further complications and injustices. But as long as these high stakes examinations are the only measure of 'competence' among young people, the system will continue to disadvantage the already disadvantaged.

As we consider the ubiquity of standards-based reforms, we need to ask how clear these standards are made to teachers and how accessible they are to students across a variety of socio-economic backgrounds. Otherwise, they just become an excuse for bashing schools that are insufficiently resourced and teachers that are inadequately supported. The second question we have got to ask is how we can improve the extent to which learner competencies associated with these reform initiatives can be measured. An even bigger question is how long we can pretend that high stakes examination systems that pervade education systems in developing countries are distributing social opportunities fairly across learners from different socio-economic backgrounds.

References

- Bernstein, B. (1996) Pedagogy, Symbolic Control and Identity: Theory, Research, Critique. London: Taylor and Francis.

- Bishop, A. J. (1988) Mathematical Enculturation: A Cultural Perspective on Mathematics Education. Dordrecht: Kluwer.
- Bond, L. (1994) Rethinking assessment and its role in supporting educational reform. North Central Regional Educational Laboratory. Available on: http://www.ncrel.org/sdrs/areas/issues/methods/assment/as700.htm
- Bourdieu, P. (1990) *The Logic of Practice*. Oxford: Blackwell.
- Broadfoot, P. (1996) Assessment and learning: Power or partnership? In H. Goldstein and T. Lewis (Eds.) *Assessment: Problems, developments and statistical issues*. Chichester: John Wiley & Sons Ltd.
- Blum, W. and Niss, M. (1991) 'Applied Mathematical Problem Solving, Modelling, Applications, and Links to Other Subjects – State, Trends and Issues in Mathematics Instruction'. In *Educational Studies in Mathematics* **22**, 37 – 68.
- Cockcroft, W. (1982) Mathematics Counts: Report into Inquiry into the Teaching of Mathematics in Schools. London: HMSO.
- Cooper, B. and Dunne, M. (1998) 'Anyone for tennis? Social class differences in children's responses to national curriculum mathematics testing'. In *The Sociological Review*.
- Cooper, B. and Dunne, M. (2000) Assessing Children's Mathematical Knowledge: Social class, sex and problem-solving. Open University Press. Buckingham.
- Darling-Hammond, L. (1994) 'Performance-based assessment and educational equity'. In *Harvard Educational Review*. Vol. 64, 5 – 29.
- Dowling, P. (1998) The Sociology of Mathematics Education: Mathematical Myths/Pedagogic Texts. The Falmer Press: London and Washington, D.C.
- Edwards, D. & Mercer, N. (1987) Common knowledge: the development of understanding in the classroom. London: Methuen.
- Gipps, C. (1996) 'Assessment for the millennium: Form, function and feedback'. Inaugural professional lecture delivered at the Institute of Education, University of London on June 6, 1996. Institute of Education, University of London.
- Hill, P. (1994) 'Putting the national profiles to use'. In *Unicorn*, Vol. 20, No. 2, 36 – 42.
- Jansen, J. (2002) 'Can research inform education policy in developing countries? A South African experience'. In *International Journal of Education Development*, 23 (1), 647 – 664.
- Kellaghan, T. & Greaney, V. (1992) Using Examinations to Improve Education: A Study in Fourteen African Countries. World Bank Technical Paper number 165.
- Kokome, N. (1990) The teaching of mathematics in Lesotho secondary schools: A package of ideas. Unpublished Diploma in Advanced Studies in Education Dissertation. University of Bristol.
- Morrow, W. (1992) A picture holds us captive. In Pendlebury, S., Hudson, L., Shalem, Y., Bensusan, D. (eds.) Kenton at Broederstroem: Conference proceedings.
- Moru, K. (1994) Developing the quality of pupil learning in secondary mathematics in Lesotho. Unpublished M ED Dissertation. University of Bristol.
- Nyabanyaba, T. (2002). Examining Examination: The ordinary level (O level) mathematics examination in Lesotho and the impact of recent trends on Basotho students' epistemological access. Unpublished Doctoral Thesis. University of the Witwatersrand. Johannesburg.
- Pennycuick, D. (1993) School effectiveness in developing countries – A summary of the research evidence. Education Research Paper. No. 1. Available on: www.dfid.gov.uk/pubs/files/schooleffectdevcountriesedpaper01.pdf
- Pule, P. (1991) Examinations Council of Lesotho: Annual Report. April 1998 to March 1991. August 1991. Maseru.

- Reville. P. (2002) Making standards work. HGSE news. Available on the Harvard Graduate School of Education website: http://www.gse.harvard.edu/news/features/reville03212002.html
- Rowie, K. (2000) 'Assessment, league tables and school effectiveness: Consider the issues and 'Let's get real'!' In the *Journal of Educational Enquiry*, 1(1), 73 – 98.

Determining Standards

Moderation of marking with specific reference to external moderation performed by Umalusi as a Quality Assurance Council

Prince Masilo, Umalusi

Abstract

Education standards can be an important tool in improving learners' achievement, because they inform learners and their parents about what society considers essential knowledge to be acquired during a given education cycle. They also provide taxpayers with benchmarks to judge how well public schools are performing. Employers look at standards to get employees with skills that are relevant to the various sectors of the economy.

It is a challenge to determine standards of knowledge, skill and competence which are relevant and responsive to the economic demands of the country, and which will have national and international status, providing clear employment, progression and personal development opportunities for learners. Umalusi and the national Department of Education have joint responsibility in providing standards for good assessment for schools, vocational education and training, and adult education and training.

This paper will discuss the processes and procedures employed by Umalusi in ensuring that marking is as valid and reliable as possible, thus ensuring that standards are maintained. The paper will specifically discuss issues relating to the standards in marking in the Senior Certificate. Umalusi's quality assurance systems for moderation of marking will be highlighted.

Moderation of marking is very crucial so as to ensure the following:

- That standards are aligned within and across centres,
- That the marks scheme differentiates sufficiently between the range of responses, and
- That there is consistency with marking or moderation of comparable units.

This moderation is quite often done against specified criteria.

This paper will then proceed to discuss the processes and procedures employed by Umalusi in ensuring that moderation of marking of the Senior Certificate Examinations, is valid, reliable and fair. The paper will highlight criteria used by Umalusi's external moderators, reporting that emanates from moderation of marking, and finally highlight challenges that Umalusi has experienced with regard to moderation of marking.

Introduction

Research indicates that a variety of factors affect marking reliability. Hartog and Rhodes (1935) found that, compared with impression marking, analytical marking slightly reduced the variation of marks awarded by different markers. Analytical marking involves allocating certain proportions of the marks for different qualities. Impression marking is based upon a general impression of the essay. Wiseman and Wrigley (1958) found that examiners' leniency was affected by the question, but this was a much smaller effect than the candidates' answers. Later Black (1962) found that some examiners' leniency was affected by the task tackled by the candidates in the examination.

In English as a Foreign Language examinations there is evidence that other factors affect marking. Training can bring examiners' differences in leniency (inter-rater reliability) to an acceptable level but it cannot eliminate them. The training can improve the consistency of examiners' individual marking (intra-rater reliability: Lunz *et al.*, 1990; Stahl & Lunz, 1991; Weigle, 1998). There is evidence that the leniency of individual examiners remains consistent, despite re-training (Lunz & O'Neill, 1997).

Coffman and Kurfman (1968) have shown that marking behaviour of a single rater changes over the marking period, although Lunz and Stahl (1990) found that markers' leniency is reasonably consistent over time. Recently, Pinot de Moira et al. (2002) found that for English, markers' leniency decreased slightly throughout the marking period. Spear (1997) found that good work tended to be assessed more favourably when it followed work of a lower standard than when it preceded such work.

Wolf (1995) argued that assessor networks or discussion between examiners is needed for reliability. She drew on work by Black *et al.* (1989). They reported that, for the communication module in the Scottish National Certificate, the assessors had found it difficult to interpret the criteria. They therefore founded a network where standards were discussed, which led to a common understanding of the criteria and facilitated an improvement in reliability, especially the inter-rater reliability.

Strong (1995) summarises some of the main points that have been discussed above as follows. In terms of the latter two types of reliability (inter-rate and intra-rater), there is overwhelming evidence that the scoring is very unreliable unless certain procedures are followed. These procedures include:

1. setting the scoring criteria in advance;
2. providing samples answers for the markers;
3. training the markers to use the criteria;

4. scoring each paper twice, and a third time if there is too much difference in the scores attributed to the same paper.

Umalusi's quality assurance of marking—processes and procedures

Moderation of marking is one of the processes utilized by Umalusi to ensure that marking is conducted in accordance with agreed practices and standards in order to ensure validity, reliability and practicability of processes, as well as to ensure that national standards are applied uniformly.

Moderation of marking takes place using three approaches:

- Approval of memorandum discussion decisions;
- Centralised moderation of marking; and
- On-site moderation of marking.

Approval of the pre-marking/memorandum discussion decisions

A week before the actual marking takes place or a day after an examination has been written, the Chief Marker, Internal Moderator, and Senior Markers meet to discuss the memorandum. Sample scripts are used as a way of standardizing the memorandum. Sufficient time is allocated to this very important exercise.

On the first day of marking quality time is devoted to the discussion of the memorandum with all Markers. Markers come prepared with question papers and suggested answers. This process takes place the whole of the first day. A sample of scripts is prepared by the Chief Marker as a way of training all Markers, with a view to finalizing the memorandum. Changes/additions to the marking memorandum, with legitimate motivations, are agreed upon and approved by the Chief Marker and Internal Moderator.

Centralised moderation of marking

This is a process in which the nationally examined subjects are moderated. Assessment bodies submit a sample of scripts to Umalusi's offices in Pretoria following the commencement of marking. A panel of moderators for each of the subjects is located at Umalusi offices during the marking process. They complete the moderation of scripts within 48 hours upon receipt thereof. Reports are sent to assessment bodies on a daily basis, in which recommendations for improving marking are made to assessment bodies.

On-site moderation of marking

Moderators are deployed to marking centres across all assessment bodies to moderate a sample of 20 HG and 20 SG marked scripts. Recommendations are provided to the Chief Marker and/or Internal Moderator, who in turn ensure that the recommendations are implemented.

A critical evaluation of Umalusi's process and procedures

The question is whether these processes are effective. Do they ensure quality assurance? Which ones are more effective and what makes them effective?

Approval of the pre-marking/memorandum discussion decisions

For the nationally examined papers this process has proved to be successful and effective in that the discussion is guided by the external moderators, although chaired by the internal moderator. On the second day the memorandum is finalized, translated, and signed-off by the external moderator. All marking centres in the country use the same standardized marking memorandum.

For papers set by assessment bodies, only the centres visited by deployed external moderator receive support and guidance. All other centres rely on their internal moderators and chief markers in case of a need to add/alter the memorandum. Nevertheless the intervention has proven to be effective and efficient.

For both process the moderators ensured that the final memorandum is appropriate, does not unfairly/fairly advantage candidates, and allows for various alternatives where possible.

Centralised moderation of marking

The scripts submitted as a sample of marked scripts as soon as they became available are moderated within the first week of the marking session. Immediate feedback is provided to assessment bodies on the quality of marking and, concomitantly, recommendations on the improvement of marking.

The appropriateness of the standard and quality of marking across all public assessment bodies is ensured. Generally markers adhere to the memorandum as agreed upon at the central memorandum discussion. There is a remarkable consistency in the marks allocated.

On-site moderation of marking

For this process the following impact has been observed:

- markers are appointed in terms of the Personnel Administrative Measures (PAM) document;
- consensus is reached on memorandum interpretation;
- relevant additions or alternatives to be included in the memorandum are accepted and finalized;
- all markers have a clear understanding of the memorandum in order to ensure accurate and consistent allocation of marks to candidates;
- inter-rater reliability has been maximized;
- internal moderation process are effective and efficient in maintaining appropriate standards in marking;
- immediate feedback is provided during the marking process so that learners are not unfairly advantaged/disadvantaged.

It is generally observed and **acknowledged** that the intervention by Umalusi at the marking centres is fruitful in that:

- The memo discussions are constructive with the chief markers participating freely, and in good spirit.
- The chief markers are conscientious about sticking to the memorandum during marking.

- It is clear that all assessment bodies have internal moderation procedures in place during the marking session.
- The external moderators find no evidence of any irregularities in the sample drawn from the provinces.
- The marking process is well controlled.
- The chief marker is supported immediately by the external moderator on issues of marking.
- Candidates' responses are included as alternative answers to the memorandum.
- Accuracy in the calculation and recording of marks has improved as the chief marker became vigilant.
- Marks for each question are entered in a pencil on the cover of the answer sheet and re-written in a red pen only after being verified by the EA and the Chief marker.

Conclusion

Looking back at what has been treated in this paper so far about marking reliability, as an educational leader you ought to be saying something such as, "All right, this measurement business about consistency is all well and good, but what does it have to do with the real world of what goes on in schools?" It's a good question, and the answer is "often, not much." In short, educational leaders need to know what reliability is, and that it comes in several forms.

When educators think of reliability, they ought to automatically think of its synonym, namely, consistency. Assessment reliability refers to consistency of measurement. In everyday language, if an assessment device is reliable, it measures consistency. One of the purposes of any systematic method of assigning marks is to establish greater uniformity among markers in their marking practices, and hence in the meaning of the marks they issue. Marks will tend to lose their meaning if the marking processes and procedures lack a clearly defined marking system or does not require markers to mark in conformity with the system. The extent to which the marking system satisfactorily reflects these considerations will largely determine its success or failure.

Umalusi regards marking reliability as a very important aspect of Senior Certificate. It endeavours to ensure that through its moderation of marking process consistency and reliability become key. This is done through quality procedures to standardize marking, which include:

- Memorandum discussion meetings;
- Marking of sample scripts;
- Moderation of marking;
- Training of markers; and
- Finalization of the memorandum.

These processes have proved to be effective. However more needs to be done in order to ensure marker reliability. Umalusi needs to streamline and strengthen the deployment of external moderators to nearly a large sample of centres. At the moment subjects with a

considerable number of candidates are chosen for moderation of marking. This makes it difficult to get sufficient information about smaller subjects. Although Umalusi relies on internal moderators and chief markers to do the quality work, that leaves much to be desired, as seen from their reports.

Training of markers is a key lever of impact in marker reliability. Unfortunately this process does not take place uniformly in all assessment bodies. This is a challenge for Umalusi. This challenge is taken up by the Quality Assurance of Assessment unit at Umalusi and there are improvements on yearly basis. Hence there is stability in the system and improvement in candidates' results.

References

- Black, J.H., Hall, J., Martin, S. & Yates, J. (1989) *The quality of assessment: case studies in the National Certificate* (Scotish Council for Research in Education).
- Coffman, W. & Kurfman, D. (1968) 'A comparison of two methods of reading essay examinations'. In *American Education Research Journal*, 5(1), 11-120.
- Hartog, P & Rhodes, E.C. (1935) 'The marks of examiners, in: E.L. Black, (1962) The marking of GCE scripts'. In *British Journal of Education Studies*, 11, 61-71
- Lunz, M.E. & O'Neill, T.R. (1997) 'A longitudinal study of judge leniency and consistency'. A paper presented at the *Annual Meeting of the American Educational Research Association*, Chicago, 24-28 March.
- Lunz, M.E. & Stahl, J.A. (1990) 'Judge consistency and severity across grading periods'. In *Evaluation and the Health Professionals*, 13(4), 425-444.
- Pinot de Moira, A., Massey, C., Baird, J. & Morrissy, M. (2002) 'Marking consistency over time'. In *Research in Education*, 67, 79-87
- Spear, M. (1997) 'The influence of contrast effects upon teachers' marks'. In *Educational Research*, 39(2), 229-233.
- Stahl, J.A. & Lunz, M.E. (1991) 'Judge performance reports: media and message'. A paper presented at the *Annual Meeting of the American Educational Research Association*, San Francisco, CA, 3-7 April.
- Strong, G. (1995) 'A survey of issues and item writing in language testing'. In *Thought Currents in English Literature*, 68, 281-312.
- Weigle, S. (1998) 'Using FACETS to model rater training effects.' In *Language Testing*, 15(2), 263-287.

Using the internet to monitor the moderation of School Based Assessment (SBA)

A first step to innovation of SBA

Clive Long, IEB

Abstract

With the current introduction of the National Curriculum Statement the potential of SBA for improving the teaching and learning process by providing feedback to learners about their learning, to the teachers about their teaching and educational leaders about the model of outcomes based assessment (OBA) they are implementing, is excellent. To achieve this potential, however, assessment bodies need to signal the great importance they attach to effective and innovative SBA.

As a first step, the IEB believes it must ensure, and be seen to ensure, the participation of all schools, teachers and subjects in the process by effectively monitoring the moderation processes at all levels. To this end it is implementing an internet based system that reduces its administrative load, is transparent, allows the assessment agency to monitor the clustering process as it happens and produce reports that can alert school management to problems while there is still time to take corrective action.

As a second step, the IEB is using its sample moderation and feedback processes as well as its regional user group structure to help ensure that SBA becomes more innovative and contributes to the debate as to how the question papers of 2008 use assessment standards to assess the extent to which candidates achieve outcomes and enable reporting against levels. This paper describes how the IEB

- uses the internet to monitor cluster moderation;
- integrates the internet based process with the moderation of SBA at the various levels; and
- is encouraging teachers of physical science to move their School Based Assessment in the direction of the assessment of learning outcomes.

Introduction

Those who believe in the value of SBA question whether the external examinations assess what proponents claim. When teacher performance is judged by external examination results many teachers, hardly surprisingly, teach to the test. This can be the start of a downward spiral. The answers to questions from old examination papers increasingly define the curriculum. To be able to answer questions in the examination, defines the aim of teaching. Rote learning and the drill of algorithms becomes the practice. Even the more demanding questions designed to assess the higher order cognitive skills can become a better indicator of the number of times candidates have practised that type of question than of their cognitive abilities. In short, external assessments of this sort

- produce results that are an indicator of the candidate's ability to answer the examination rather than of the candidate's ability; and
- sacrifice any attempt at "assessment for learning" in favour of reproducing the correct answers and good results.

Those who believe SBA is of little value put forward many arguments to support their position. These arguments are important and need to be addressed if teachers are to give SBA the attention educational authorities demand:

- Many regard SBA as unreliable.
- Critics of SBA invariably raise the problem of a SBA task's authenticity. Is it the work of the candidate or did the candidate get help, or in this day and age, did the candidate simply download it from the internet?
- SBA
    - is time consuming for the candidate
    - increases the teacher's load
    - generates excessive paper work as everyone with responsibility for the implementation of SBA moves to try ensure it is validity and reliability.
- Authorities signal the lack of importance they attach to SBA by
    - statistically moderating SBA against the external examination
    - increasing the amount of standardised testing to ensure accountability
    - collecting a single SBA mark which many contend could be more efficiently done using a test.

In spite of the very real difficulties with SBA there are good reasons to retain or even increase the quantity. With the introduction of the National Curriculum Statement, to be assessed at grade 12 in 2008 for the first time, the South African National Department of Education has recommitted itself to a role for SBA.

In this case this commitment seems related to the growing world wide realisation, based on research evidence, that formative assessment for learning improves the quality of learning. Dylan Wiliam writes, referring to the 'spiral' described in the first paragraph above,

> In place of this vicious spiral, I propose developing a system of summative assessment based on moderated teacher assessment. A separate system, relying on

> 'light sampling' of the performance of schools would provide stable and robust information for the purpose of accountability and policy-information. (Wiliam, 2001)

From the above introductory comments it should be clear that SBA has the potential to offer several important educational benefits. However for any system of SBA to receive general support from teachers several issues need to be addressed.

- SBA must be seen to be important, it must be given as much status as possible.
- Measures need to be introduced to help teachers
    - see educational value in SBA and motivate them to commit themselves to it
    - during the change process
    - by streamlining the administrative processes.
- For authorities to use SBA measures need to be in place to ensure SBA is as reliable as possible.

SBA must be seen to be important

Documents describing policy and processes for both the quality assurance body and the assessment agency must be available to clients (schools). The quality assurance body and the assessment agency must monitor and be seen to monitor the implementation of the aspects of policy and processes for which each is responsible. The assessment body must take punitive action as described by its own policy (within the framework of the policy of the quality assurance body) if an institution does not meet any of the SBA requirements. In order to do this the assessment agency needs to have absolute confidence in its processes.

Clients must have detailed knowledge and a complete understanding of processes like the statistical adjustment of SBA. A lack of knowledge and understanding of the process is often the cause of discontent. Currently the average for the SBA achieved by a group of candidates at a particular institution has to be more than 15% above the average the group achieves in the examination before any downward statistical adjustment is made. There is also a check using the standard deviation to ensure that the spread of marks in the SBA is in line with the spread in the examination. In addition any institution that sends in SBA results with an average below that which the group achieves in the final examination is adjusted upwards so that the group is not disadvantaged. There is ample space for candidates who excel at SBA to gain an advantage and those that perform poorly to be penalised by this system. Teachers need to understand this and exercise great responsibility with the implementation of SBA.

However there are teachers that are negative about SBA simply because there is statistical moderation. Some of these teachers express the view that SBA is a waste of time. It seems that this negative sentiment could be negated to some extent if there was an incentive to be more conscientious about SBA. Suppose the SBA results of the candidates of a particular teacher never required statistical adjustment over an extended period, say 3 years. Would it not be a good idea to allocate special status to such teachers and accept the results of their learners without adjustment? The results can still be monitored statistically as before. If the results are out of line they should be investigated the following year but stand unaltered for the current year. If a satisfactory explanation is forthcoming the teachers' status is retained. If the discrepancy between the SBA average and the examination average cannot be

explained the teachers lose their status and the statistical moderation becomes operative at the end of that year. This would send a signal that the authorities do really value SBA and wish to reward teachers and learners that implement it conscientiously.

Measures need to be introduced to motivate teacher support for SBA

In an examination driven system it is difficult to persuade teachers that SBA has much value. The teacher is judged by learners, school management and parents against the results the teacher's learners achieve in the final examination. The driving force is good results.

Assessment for learning

In June 2005 groups of educators were called together by the National Department of Education (DoE) to draft the subject assessment guidelines for the new South African National Curriculum Statement (NCS). Amongst the documentation that was supplied to the educators was the article "Inside the black box: Raising standards through classroom assessment" (Black and Wiliam, 1998). The same document was distributed at the Gauteng Department of Education's (GDE) "Assessment for Learning Conference" held on 5 and 6 May 2006. The IEB National Assessment Conference held in August 2005 also promoted "Assessment for Learning". The fact that the "assessment for learning" approach is now regarded as important is further illustrated by the public report put out by Umalusi in which it criticises teachers, by implication, for using SBA to prepare learners for the final examination, instead of setting assessment tasks that promote learning!

The following was noted in Accounting portfolios …

- There was very little creativity shown and most class tests are exercises out of books or old papers. This means that it is just a repetition and the learner's context is not considered. Tests and examinations are to a large extent a mere regurgitation of old question papers.…
- Assessment tasks set … lacked higher order questions. (Umalusi, 2005 p33).

The message now seems clear. On this evidence Umalusi expects SBA to be used to improve learning and shift the emphasis away from producing good examination results.

Use summative tasks formatively

Though "Inside the black box" (Black and Wiliam, 1998) focuses on the use of formative assessment to promote learning summative assessment can also be used formatively. In the South African context to use summative assessment formatively will be very important to moving in the direction of assessing for learning for several reasons:

- An examination of the subject assessment guidelines still shows a heavy emphasis on summative assessment;
- SBA is moderated against the final examination; and
- Teachers understandably will still be teaching to the test.

The Western Cape Education department is already collecting marks from the final external examination per question in a way that can help use the examination results formatively.

The IEB did the same thing at a pilot project level at the end of 2005 and will be extending the project in the future. However "assessment for learning" goes a great deal further than this. Assessment for learning promotes an interaction between teacher and learner in which the teacher sets tasks that create opportunities to gather information about the learner's learning. The teacher then needs to use the learners' response to the task to help the learners think metacognitively about their own learning. They need to identify what they know, what they need to learn to achieve their learning goal and then be able to strategise as to how best to get to their goal. The process simultaneously informs the teaching process and can be used to strengthen the curriculum itself.

Set summative tasks for learning

The critical issue to using summative assessment formatively will be the type assessment tasks that are set. These tasks need to create the opportunities referred to above that reveal what the learners know and can do, rather than simply be a mechanism for allocating marks. Teachers need to learn how to set these tasks and how to assess them in a way that they can use to give feedback to learners about their learning. They also need to find ways to reduce the emphasis on the result. Fortunately the NCS with its learning outcomes and assessment standards provides criteria against which learning can be assessed. This is a place for teachers to start.

Give feedback

Not only will teachers who assess for learning need to give feedback to their learners, the IEB needs to give feedback to teachers. See Appendix A for a flowchart of the processes the IEB uses to monitor and moderate SBA and give written feedback to teachers after the sample moderation process at the end of the academic year.

In spite of the current hostile environment with its emphasis on results the IEB is, through the SBA process, promoting alternative assessment strategies more in line with assessing for learning. The science report back forms include a checklist (see Appendix B) that will be used at both the cluster moderation process and during the sample moderation process this year. (These forms update similar forms used previously.) A version of the completed form is returned to the school in January of the following year. The portfolio moderator also reports her findings at the sample moderation process during the National User Group Conference in February. This is attended by some 180 teachers from the 160 schools that enter for the senior certificate with the IEB. At this time there is opportunity to discuss issues from the previous year. The User Group conferences are also used to work through exemplar material that promotes better learning, teaching and assessment.

Using the internet to ensure SBA is reliable and streamlines administration

The first step to reliability is to ensure that all schools, all teachers and all candidates involved are doing what they are required to do.

The IEB

In the South African State system half a million candidates are assessed in a final written examination in all subjects across all provincial examining bodies each year. In contrast there are some 7000 candidates in schools that write the IEB examinations (IEB schools).

Like those of provincial departments these schools are very different. There are schools that are

- extremely well resourced and others that are very poorly resourced;
- run as businesses and others like those that are affiliated to different religious groupings, like Christians (Catholics, Methodists, Anglicans and several other smaller groups), Jews and Muslims, that are set up to serve their communities;
- in urban areas and others that are in rural areas and isolated from other IEB schools;
- in different countries (Namibia, Swaziland, Mozambique);
- of medium size (100 grade 12 candidates) and those that are small (less than ten grade 12 candidates).

The relatively small number of IEB schools means that the IEB does not have a provincial, regional or district structure. Add to this the heterogeneous nature of the IEB schools and it becomes apparent that the IEB is faced with a unique set of problems when it comes to monitoring and moderating SBA.

However, because IEB schools are spread across Southern Africa, of necessity, the IEB and its schools have increasingly begun to rely on electronic methods of communication, like email and the internet. With the exception of literally one or two small schools that still rely on faxes the IEB is in email contact with its schools. Virtually all also have access to the internet. Electronic communication is then the one way the IEB can monitor SBA and reduce the administrative load on teachers and the IEB.

The Cluster Report Manager (CRM) Program

Overview

This is a computer program hosted on the IEB website (www.ieb.co.za). Anyone can access it by clicking on the link on the IEB home page. The program interfaces with a database of information relevant to clusters, cluster meetings and the monitoring of SBA.

Teachers

Cluster members (teachers) can go onto www.ieb.co.za and elect to view

- The names of the schools with which the school is grouped within a cluster for a subject or for all the school's subjects for grade 12 and separately for grade 9
- The name, school and contact details of the cluster leader
- Details of the next cluster meeting: date, time, venue, and purpose. They can also check to see who has indicated they will be attending and who has apologised.

Teachers are also expected to respond to notices of meetings that they receive by email. They have to click on the link in the email while on line and indicate whether they will be attending or tender an apology.

Cluster Leaders

Cluster leaders are elected at the first meeting of the year which is scheduled to be held before the end of March. The outgoing cluster leader enters the details of the newly elected cluster leader using the CRM. The CRM automatically replaces the details of the old cluster

leader with that of the new cluster leader and generates a password which it emails to the new cluster leader. The password gives the new cluster leader access to edit and input certain data and send out notices of meetings using the CRM.

It is clearly important that cluster leaders are confident using computers and have convenient access to the internet. Cluster leaders have several responsibilities with regard to the CRM. Cluster leaders need to

- Keep the contact details of the teachers in their cluster up dated on the CRM;
- Send out notices of meetings using the CRM;
- Monitor the response to notices of meetings using the CRM and take action if problems are identified;
- Edit the data on www.ieb.co.za to reflect actual attendance at meetings;
- Record on the CRM that cluster moderation reports have been submitted to the IEB, by fax or post, after the final cluster meeting of the year;
- Enter the name, school and contact details of the newly elected cluster and submit to the IEB on the CRM at the end of their term of office.

IEB assessment specialists

IEB assessment specialists are permanent professional IEB staff members that play a leadership role in the professional development of internal moderators, examiners and teachers in the specific subjects for which they are responsible. They are also required to keep abreast of the latest trends in the assessment of their subjects nationally and internationally and motivate for change. With regard to the CRM they are responsible for

- ensuring the lists of schools on the CRM reflect schools currently registered with the IEB;
- ensuring the CRM reflects the current list of subjects schools offer;
- answer queries from, and give support to, teachers struggling to use the CRM, i.e. supply passwords when the previous cluster leader has left an IEB school without registering a new cluster leader or holding a meeting;
- generating and distributing reports using the CRM, i.e. send emails to the principals of all schools identified as having not attended a cluster meeting for any subject. Such reports are to inform principals of the situation and urge them to take corrective action. In this way problems can be solved before the final cluster meetings and the submission of portfolios for sample moderation;
- supply the sample moderation teams at the end of the year with reports from the CRM to inform the moderation committees of the attendance history of schools and the cluster groups to which the schools belong.

IEB IT Consultant

This expert is available to:

- maintain the CRM Program;
- change or adapt the program when required;
- support the maintenance of the database in special cases.

Advantages

Database maintenance

The maintenance of the database is done by the people using the CRM. The cluster leader enters email address of cluster members once. After that the data only has to be updated as required by the person using it. When a cluster leader sends out a notice she/he does no more than before. The notice arrives on the desk of each teacher that has an email address. The member replies to the notice with the click of a mouse. The cluster leader enters the details of the new cluster leader after his/her term has expired.

Transparency

All the information entered is immediately available to the members of the cluster, the cluster leader and the IEB, whether it is

- a notice of a meeting,
- an attendance list,
- the notification of the submission of cluster reports or
- the contact details of the new cluster leader.

Errors can be noted by the people affected and they can themselves take steps to correct them.

If the IEB moves a school from one cluster to another or adds a new school the information is available to all who take the trouble to look for it. A new school has immediately available the contact details of all the cluster leaders under which it falls.

Monitoring

The IEB is able to monitor cluster meetings, attendance of schools at cluster meetings and the submission of cluster moderation reports at the end of the year by looking at the information on the website using the CRM. Alternatively the operator can make printouts of reports.

Intervention

It is possible to identify problems when there is still time to initiate corrective action. The IEB can identify problems and direct timeous intervention using a single email that is sent to all sites where there is the same problem.

New Developments

Though the current CRM was not designed for the distribution of minutes it can be used for this purpose with a little effort. However the IEB is delighted that some teachers are now engaging with the CRM to the extent that they are suggesting improvements. Modifying the CRM to distribute and possibly store minutes has a number of advantages and adapting the CRM for these purposes are under investigation. Clearly this would have the advantage of further reducing paper work.

Problems

The IEB is able to successfully monitor and moderate SBA through its sample moderating process at the end of the year (see Appendix A) in spite of the fact that the implementation of the CRM process has yet to reach the ideal described above. This also means that the IEB

and its clients have yet to experience all the benefits that the full implementation of the CRM promises. Some of the reasons why the CRM has yet to deliver all it promises include:

- The CRM is associated with SBA, a process about which many teachers are still sceptical and with which they are still trying to come to terms. SBA is frankly seen by many as a huge drain on their time and a distraction from getting on with preparing candidates for the examinations which is the really important task.
- The CRM represents another change at a time of extensive curriculum change and new challenges on all fronts. SBA was introduced only a few years ago with a monitoring process based on the submission of reports on paper. Before teachers were able to get the first process working properly a new system was introduced. The CRM is new and it is a radical departure from the processes with which teachers are familiar.
- With some justification teachers see the CRM as designed to benefit the IEB and see little benefit for them. This is beginning to change.
- While the CRM was introduced at the beginning of 2005 the new associated processes connected with the establishment of clusters were implemented for the first time only 4 months ago, at the beginning of 2006. 2006 is the first year the CRM is being fully implemented. The process is very new.
- While virtually all schools have email and internet access it has become apparent that many operate with a dial-up connection and one computer. This computer is often also used for administration by a secretary who prints email messages, deletes them and places the hard copy in the teacher's pigeon-hole or mailbox. Instructions that read "click on the link (blue text) at the bottom of this email …" do not make much sense if you are reading a printout!
- While many teachers have their own email addresses, those that do not have to collect their messages from the central school email. This makes nonsense of sending notices via the CRM to each cluster member's desk and allowing a response with the click of the mouse.
- The level of computer literacy amongst teachers at client schools is less than was anticipated. There are teachers that have sought help form the IEB who have passed remarks that make it clear that they lack confidence working on the internet and many suffer from computer phobia.

Concluding remarks

Under the circumstances currently prevailing in the IEB the CRM offers the best hope of effectively monitoring SBA on a day to day basis. The IEB needs to take cognisance of teachers needs and make the system more helpful to teachers, to help those that are fearful become more confident and sell the idea of the CRM by demonstrating its advantages to teachers and school management.

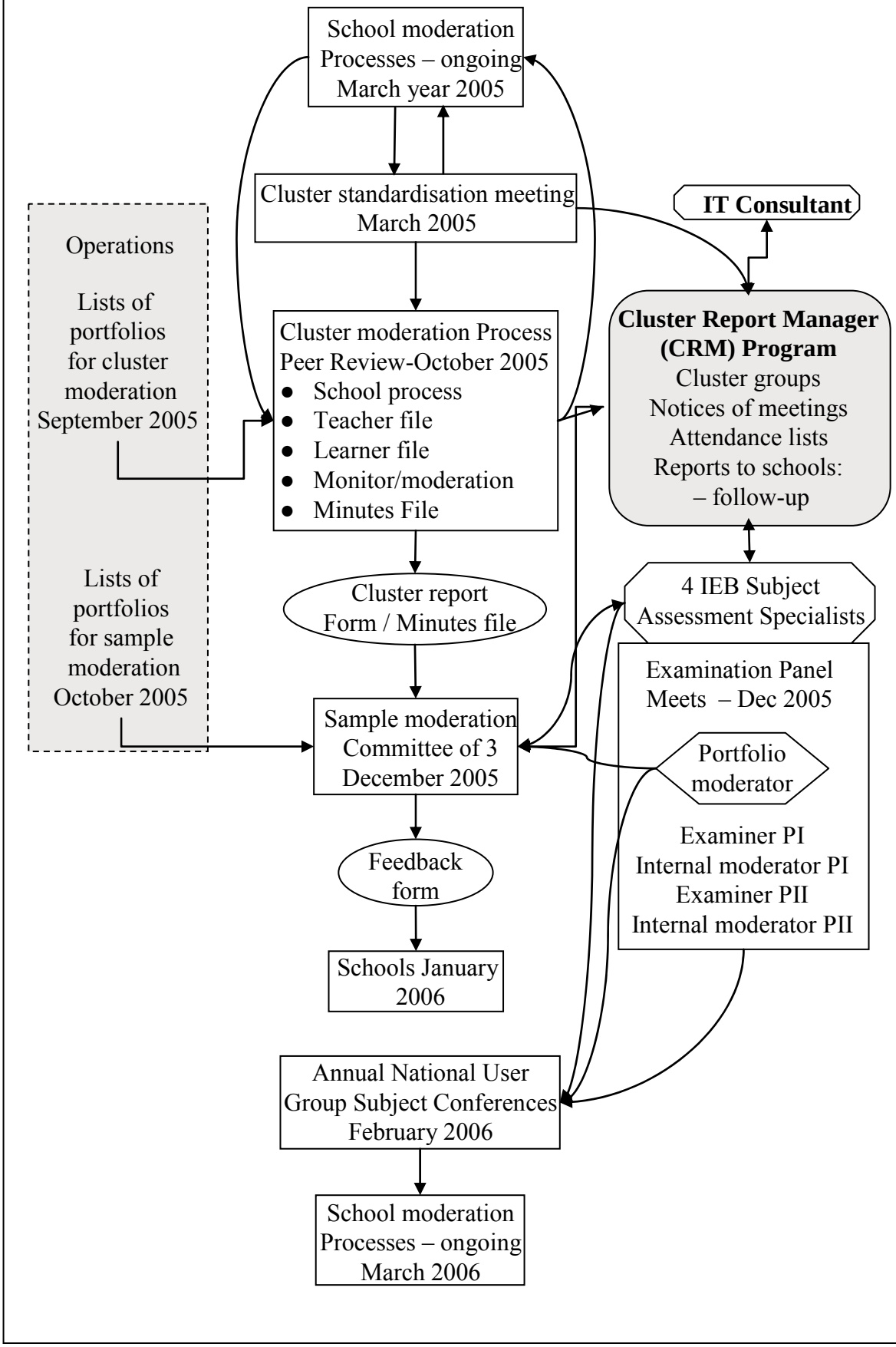
However there is a much bigger issue with regard to SBA. Teachers have to be convinced that SBA itself is worth doing and doing well. While the IEB is convinced of the value of SBA and is playing its part enthusiastically in promoting SBA, aspects of this task are beyond its control and dependent on the importance the nation attaches to SBA. This will become evident to the extent to which the nation (parents, learners, teachers, education departments

and government) looks to reduce the emphasis on examination results and place it on improving learning and raising standards of education.

References

- Black, P. and Wiliam, D. (1998) *Inside the black box Raising standards through classroom assessment*, School of Education, King's College London.
- Umalusi, (the South African quality assurer in the general and further education and training bands of the national qualifications framework (NQF)) (December 2005). Report on the Quality Assurance of the Senior Certificate Examination 2005.
- Wiliam, D. (Autumn 2001). What is wrong with our educational assessments and what can be done about it. Education Review 15(1)

A Flowchart of the monitoring and moderation processes connected with SBA



Appendix B (page 1)

CLUSTER MODERATION CHECKLIST

The Cluster Leader must consult the "IEB Manual for the Moderation of School Based Assessment for the Senior Certificate Examination" (at www.ieb.co.za) for current information concerning the use of this form and the associated processes.

SUBJECT:	**PHYSICAL SCIENCE**		
Teacher's Name		School:	
Moderator's Name		School:	

Teacher's portfolio: IEB list candidates for sample moderation is available. y/n

No. of tasks available:

Informal	4	with memo	y/n	Practical	8	with memo	y/n		
Project	0/1	with memos	y/n	Tests	4	with memo	y/n	Examinations	1 exam	with memo	y/n

Marking: According to memo? y/n — Moderator please sign the front of the teachers file plus at least 1 HG and 1 SG portfolio moderated (use a green pen)

Adjustment of marks (HG → SG) process: By calculation/ reference to SG candidates/ remarking/ other/ n.a./ ...

Correct Weighting: Chemistry/Physic? y/n | Gr11/gr12? y/n | Composite mark sheet available y/n

Evidence of school based moderation -Appendix H? y/n | other? moderated tests/ minutes /policy /checklists:

Analysis of practical work: Did the ***assessment*** of practical work include the following?

Identify phenomenon to be investigated	y/n	Formulate a hypothesis (PEE)	y/n	Identify variables	y/n	Design investigations	y/n
Manipulate apparatus	y/n	Make observations	y/n	Take measurements	y/n	Record and organise data	y/n
Draw conclusions	y/n	Explain observations	y/n	Explain science concepts			

Refer to any interesting/innovative work:

Give the title of any tasks and the description of any processes (administrative, teaching and learning, …)

and where to find these in the teacher's portfolio

Candidates' portfolios: 1st summary page supplied y/n | aggregation correct y/n

Task types:

Informal	4	Examinations	1 exam p1&p2	Tests	4	Practicals	8	Project	1/0
Learner achievement recorded	Yes / No					Appropriate aggregation?		Yes / No	

Additional Comment – general: candidates take a pride in their work; clearly enjoy science/

work is of high standard/ portfolios well organized/ marking promotes learning/ innovative and varied tasks

opportunities for individual initiative/ too little demanded/ traditional/ teacher appointed from June & no portfolio
work done till then/ unqualified teacher helping out/ teacher exceptionally supportive of colleagues in the cluster

TEACHER'S SIGNATURE: _____ **DATE:** _____

MODERATOR'S SIGNATURE: _____ **DATE:** _____

CLUSTER LEADER'S SIGNATURE: _____ **DATE:** _____

Appendix B (Page 2)

SAMPLE MODERATION GRADE 12 – PHYSICAL SCIENCE

To be completed and returned to the school

Examination Centre Number: _____ Date: December 2006

Records

	Response and comment
Summary of candidates' marks available (teacher file)	y/n accurate/ complete/ clear
Summary of assessment available (1st page of candidate's portfolio)	y/n accurate/ complete/ clear
Appropriate aggregation of marks (1st page of candidate's portfolio)	y/n accurate/ complete/ clear
All the tasks with marking memoranda are available in teacher's file	y/n mostly / several missing /
Appendix D for Physical Science available (Cluster Moderator's Checklist)	y/n
Evidence of school based moderation (Appendix H plus any other …)	y/n minutes /policy / checklists:
Candidate's portfolios supplied according to list from IEB	y/n +2 / no list/

Tasks completed and marked according to requirements

	Response and comment
Informal Assessment - 4 items – enter number available and comment	4 varied/innovative/traditional
Practical Work - 8 items – enter number available and comment	4 individual/ varied/ group
Model B chosen – 1 project – (cross – X - if yes)	y / n
4 controlled tests – enter number available and comment	4 original/ IEB questions/ new
1 examination (paper 1 & paper 2 from the same examination session)	4 original/ IEB questions/ new
Marking in accordance with the memoranda	y / n rubrics used/ other tools
Accountable adjustment of marks for candidates changing from HG to SG	y / n calculation/ cf SG candidates

Correct weighting: Physics : Chemistry :: 40-60% : 40-60-%	y / n	actual ratio if beyond limit
Correct weighting grade 11 : grade 12 :: 0-40% : 60%-100%	y / n	actual ratio if beyond limit

		Practicals	Informal	Tests & Exams
Level of difficulty of set tasks	**SOME STRAIGHT FORWARD, MANY CHALLENGING**	practical	informal	Tests & exams
	MANY STRAIGHT FORWARD, SOME CHALLENGING	practical	informal	Tests & exams
	ALL STRAIGHT FORWARD	practical	informal	Tests & exams

Additional Comments(including descriptions of any tasks or questions of particular merit):

Portfolio Moderator's signature _____ Date: 9 December 2006

Remarking of examination answer scripts

Finding a standard for quality assurance

Joe Cesare, Gauteng Department of Education and Coert Loock, University of Johannesburg

Abstract

The value of remarking of examination answer scripts is often limited to knowing how many changed symbols and *ad hoc* identification of the most glaring errors made by the initial markers.

In any marking action performed by people there will always be the occasion where the marker has to make a judgment call on whether or not to award a mark for a specific answer. To be fair towards the candidates the markers often use a system of marking per question to minimize the effect of this subjectivity and attain better overall consistency. It is a generally accepted educational principle that there may be small variations in marking by different markers, even if they use the same memorandum. However, the question still remains of when is the difference acceptable and when not.

With this paper the authors examine the changes in marks during remarking of a large number of subjects in a large-scale examination such as the Senior Certificate Examination in Gauteng over a number of years. The analysis is done in terms of changes in two-percent intervals on raw marks to determine a standard against which individual subjects can be measured. Raw marks are specifically used as to eliminate the effects of standardization. The analysis is spread over a number of years to determine patterns and eliminate once-off individual problems.

The effect of deliberate interventions introduced by the province to improve the quality of marking is shown in terms of changes in remarking patterns. The remarking of the subject Physical Science of one cycle is analyzed in detail to show the nature and frequency of common marking errors made during initial marking, with reference to the relationship between the nature of the question (multiple choice, calculations, etc.) and the type of error.

Introduction

This particular project initially started off as a search for a quality assurance tool to use during the marking process of the Senior Certificate Examination. The initial question posed was "how do we improve on the quality of marking?" Because remarking shows the "mistakes" made during the initial marking, it seemed a logical starting point.

It is interesting to note that there is very little information on the result of a very widely used concept that is within the general public domain. It is not as if remark results are classified as restricted, but rather as something one does not really want to highlight as it is indicative that the marking process may in fact not be as accurate as quality assurance bodies would like the general public to perceive. It is a characteristic of high volume-high stakes examinations that the public, the media and the politicians show a great interest in the examination, not because of the intrinsic value, but because of the potential value to promote specific agendas.

Misinterpretation of the remarking process could very easily erode the confidence of the public in an examination system. It was also quite an experience to observe the reactions of examiners and markers when we started discussing our analysis of remarking with them.

The concept of remarking

The concept of allowing remarking an assessment task is widely applied, not only at the levels of academic institutions but also in industries where formal examinations are conducted. Where remarking is a formal process, or part of a formal process, there is a policy or guidelines that governs the process. These policies usually include reference to "where the student is not satisfied with the result obtained" or words to that effect.

While the particulars of these policies are unique for every organisation or institution, the requirements to qualify can be grouped into main categories:

- There is some monetary fee attached. The fees are normally affordable and related to administrative costs and markers remuneration involved, but there are instances where there are marked differences in fee structure applied by different assessment bodies for identical examinations, e.g. for remarking Senior Certificate Gauteng Department of Education charges R70 per subject and the IEB charges R 400 per subject (GDE, 2006; IEB, 2006).
- In the majority of organisations, the fee is refundable if the remark results in a grade/symbol change (DOE, 2005; GDE, 2006). There are however institutions that do not make a refund, particularly web-based programmes (Purpletrain, 2006).
- The application should usually be accompanied by a motivation for the request (City University London, 2006) or the request be supported by a governing body (University of Papua New Guinea, 2006). More often than not there is a qualifying criterion attached, e.g. having a certain minimum score, or be within a certain range for a pass or distinction (Unisa, 2006), result in a qualitative change of the student's academic status (University of Addis Ababa, 2006) or be applicable to failed candidates (Open University of Hong Kong, 2006).
- Remarking applies only to written final examinations and not the practical or course work (Addis Ababa University, 2006; GDE, 2006; Technikon North Gauteng, 2006).

- Remarking is often one part of an appeals process, such as found with GSCE and A-levels (Teachernet, 2006).
- There is a time limit attached to an application being made.
- A different marker, usually a more senior marking official, will do the remarking.
- The candidate usually receives the higher of the normal and remarking scores (DOE, 2005, IEB 2006). There are a few institutions where the candidate will receive the remark score as final, regardless of the previous score (Technikon North Gauteng, 2006, University of New England, 2006, Addis Ababa University, 2006).
- Remarking is often a component of a results enquiry process, where the enquiry can trigger re-checking of grades, re-moderation of course work or remarking of written papers. This is the process applied for GCSE and A-levels and seems to be a widely followed model (Teachernet, 2006). This process stands in contrast to the procedure followed in South Africa, where rechecking and remarking are two distinct processes, with a viewing process and an appeal process two consequential processes that may follow a remark (DOE, 2005).
- The majority of assessment bodies do make a distinction between re-checking or verification (essentially a clerical checking of addition and computation) and remarking by a different, usually more senior, marker.

The rechecking process

The majority of assessment authorities allows for a re-checking process that is separated from the remarking process. While re-checking concentrates on clerical to ascertain if all work marked and marks added and computed correctly, the same actions are automatically performed during the remarking process. Very often a score will change after remarking, not because of a marking error but due to a clerical error that could have been picked up during a re-check. In the South African scenario the candidate has to choose either one of the two options. As the cost and effort required from the candidate is not significantly different ((R 12 for a recheck and R 70 for a remark (GDE, 2006)), many candidates opt for a remark. This tendency has major implications for the examining body in terms of time, logistics and infrastructure.

Our research shows that despite deliberate mechanisms to prevent transcription and arithmetical errors, they still occur and are responsible for a percentage of mark changes with remarking.

It seems as if there are two main "culprits" that manage to bypass to control measures such as checking adding and transfer of marks by another person than the marker. The first of these is mental fatigue. Senior certificate marking is always a high volume of work that must be completed in a short space of time under extreme pressure deadlines—normal 'human error' will definitely start to exact its toll. In a Physical Science script the number of digits a marker has to read, mentally interpret correctly, check on correct placements in formulas, check on manipulation and calculation and then allocate the correct number of marks, add up 36 subsections into nine question totals and transfer these correctly to a script cover and add the total, is a mind boggling number of mental computations. If the marker then has to

transcribe 106 marks to candidate 1606060160 and 160 marks to the next candidate number 1606060166, the chance of making a mistake becomes definite.

The second culprit, not always recognised, is the role of language when using numbers. It is a natural tendency to think in one's mother tongue when doing simple mental arithmetic. In Afrikaans the number 69 is pronounced as nege(9)-en-ses(6)tig and in English as six(6)ty-nine(9). Writing 69 as 96 thus becomes a common error (then you can still add the problem of 6 looking like 0 when the hand gets tired!)

The remarking process

One of the outstanding aspects of remarking is probably the fact that the examiner or a more senior marker is charged with the remarking of a script. This aspect could be considered to be a double-edged sword. On the one hand the marker will be able to pick up on deviations/alternatives much easier and consequently evaluate the candidate's response more accurately—the candidate may thus gain or lose due to the marker's ability.

The influence of the markers' experience and ability varies from subject to subject—the effect is more pronounced in subjects where there has to be interpretation of the candidate's response, as found in essay type questions and questions where analysis and synthesis are required. Where the response is purely stating of facts or simple calculations the marker's ability plays a much smaller role.

The advantage seems to always be with the candidate. With few exceptions qualification authorities give the candidate the benefit of obtaining a higher mark with remarking and ignore the score if the remark results in a lower score.

When results are standardized the initial "incorrect" scores are used to determine adjustments. The remark results are subject to the same adjustment process as the original scores. While the numbers involved in remarking are such that it may be insignificant, it would be interesting to see if there would have been the same adjustments had the remarked scores been used.

In order to cancel out the effect of any standardization adjustments, it was decided that for the purpose of this research the candidates' raw marks would be used.

In subjects such as Physical Science and Mathematics, there are often more than one method that can be applied to get the correct answer, particularly in questions that require analysis and synthesis. Examiners usually try to cover most of the alternatives in the memorandum. Other alternatives are readily recognised and credited by markers. Because mark allocation is usually not an all or nothing situation, it does become a problem when an alternative is partially correct—two different markers may give different marks and both be able to justify why they awarded the particular mark. The same concept can be applied in most subjects.

Marking done by educators is not always the objective exercise that educators promote it to be. This is particularly true in high profile-high stakes examinations across the world. The

very fact that the candidates performance are often used to measure the success of the educator of the institution, gives the marker a vested interest (even if he is not marking his own candidates) and they will tend to err to the advantage of the candidate—this is particularly true in remarking borderline cases. Markers will tend to search for possible additional marks and not necessarily mark with the same strictness as during the initial marking.

It is an unfortunate fact that exam results are used to measure—what you get from your measurement depends on your intended outcome. If the marks are very high, it is often said that the examiners set the paper too easy, or the markers were too lenient, or the standard of examinations is too low (and the opposite is equally true). If a remarking action results in general increases or decreases, it can only add fuel to the fire, as was seen in the UK when there was an overhaul of the grading system (Demopoulos, 2005, Tomlinson, 2002). It is for this reason that in an ideal world there shouldn't be significant changes with remarking—which relates directly to the topic finding a standard for quality assurance.

The viewing and appeal process in relation to remarking

It is interesting to note that there are two schools of thought on allowing a candidate access to information (in this case the actual marked script). There are instances that allow access to original scripts under supervision on request (UP, 2006; DEFS, 2006; DOE, 2005) or allow photocopies to be accessed (DEFS, 2006), while in most this is not the case.

While access to the script would probably have a significant effect on the reduction of the number of remark requests, as shown by the high percentage of scripts that have no change in marks after remarking, one of the preventative issues in allowing open access to candidates would be the logistics involved in maintaining the integrity and security of the scripts in large scale examinations.

Why use remarking for quality assurance

There are basically three reasons why candidates apply for remarking:

- Candidates on the verge of a higher symbol hope for a mark or two to get into the next range.
- Candidates require higher symbols and/or scores to qualify for admission to tertiary institutions or specific faculties and for bursaries and financial assistance.
- Candidates, their parents and educators do not trust the quality and consistency of the marking or the abilities of the markers and hence apply for a remark.

The whole concept of remarking is based on providing an opportunity to a candidate to have a script remarked and rechecked—unfortunately it has lately turned into a second chance for candidates that need higher symbols and higher scores for bursaries, etc.

In any marking action performed by people there will always be the occasion where the marker has to make a judgment call on whether to award or not to award a mark for a specific answer. The effect of this subjectivity is largely cancelled by using a system of

marking per question so that no individual is solely responsible for the final result. This gives better overall consistency and is considered to fair towards the candidates.

One of the aims and objectives of the marking is to ensure that our clients perceive our marking process as accurate, fair and consistent. With the high profile of the Senior Certificate Examination, such perceptions have a spill over to education in general. It is a generally accepted educational principle that there may be small variations in marking by different markers, even if they use the same memorandum. Our aim is to **reduce the margin of error to acceptable levels**. The remarking process can be seen as a test for our quality assurance mechanisms used in marking.

In the past our analysis of Remark/Recheck results only indicated the number of remark applications that result in a symbol change and as this could be indicative of either very small changes or of major changes in the raw mark. In terms of our processes, practices and operations this did not add any value, as we could not determine where we are going wrong. This necessitated that our approach to remark analysis had to change. The statistics provided by the remark markers were also not always that accurate and could not be used to address shortcomings in the marking process.

Analysis of the Gauteng Senior Certificate remarking results

The challenge posed to the Marking Unit was to assure the quality of marking. Quality is an aspect that is difficult to measure in education, as we do not have a rigid and objective baseline from which to evaluate. As the baseline should not be influenced by activities outside the marking process, our departure point was to analyse remark results in terms of raw marks obtained, as to eliminate the influence of standardization on the final symbol awarded. The very nature of the marking task, coupled to the approximately 1.2 million scripts marked by 6,500 people, made using observation against checklists (as it is used during monitoring) an exercise that eases the conscience rather than add value. It is just too easy to 'hide' bad practice from a monitor that is not necessarily a subject expert.

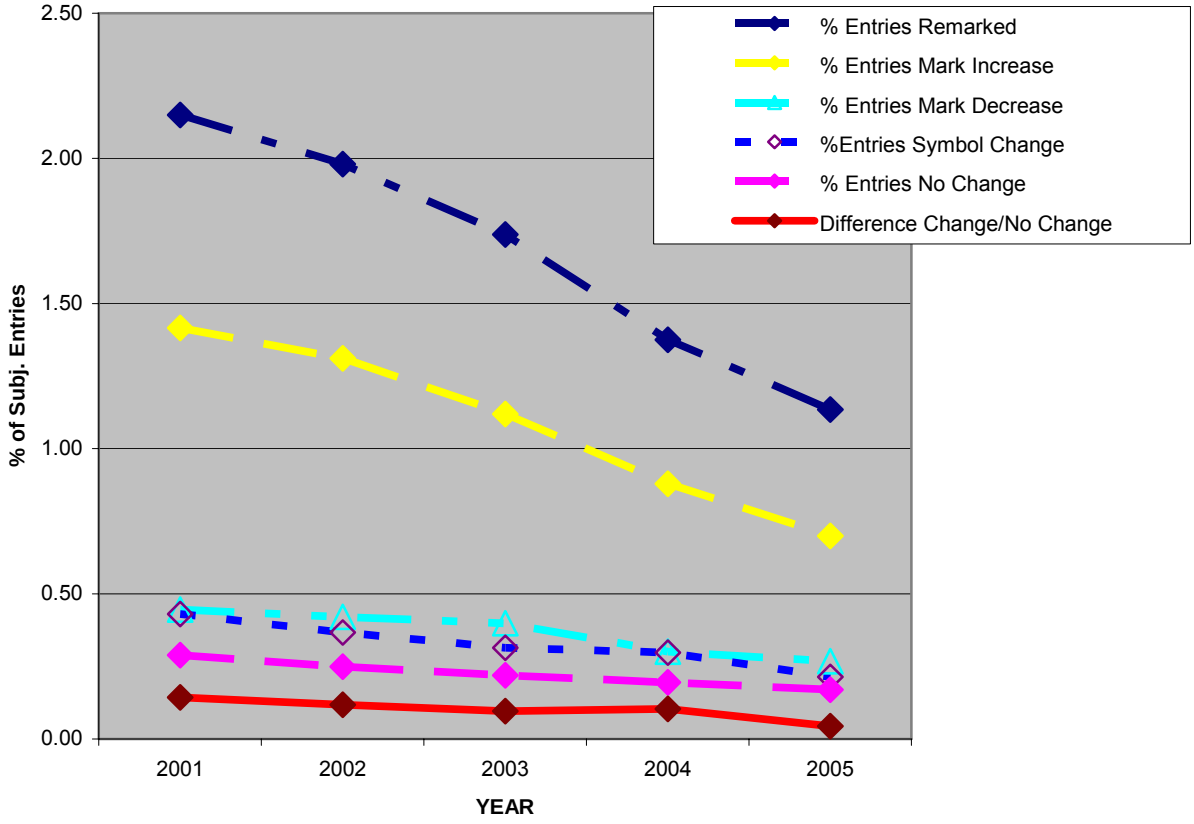
Using remarking as a vehicle to drive a quality assurance process seemed to be an option. We were thrilled when our initiative started showing positive results.

Table 1: Trends in Remarking

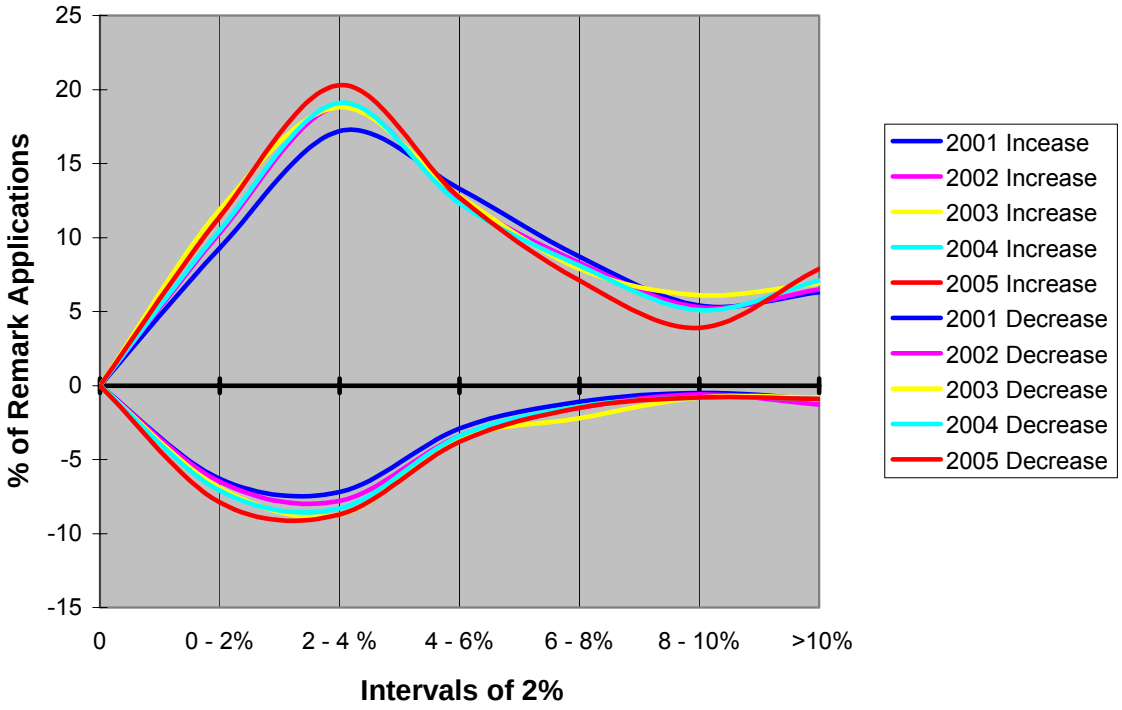
Year	Subj. Entries	Remark Applications	% Entries Remarked	%Entries No Change	%Entries Increase	%Entries Decrease	%Entries Symbol Change	Difference Change-No Change
2001	619488	13319	2.15	0.29	1.42	0.45	0.43	0.14
2002	621122	12301	1.98	0.25	1.31	0.42	0.37	0.12
2003	637904	11082	1.74	0.22	1.12	0.40	0.31	0.10
2004	664455	9136	1.37	0.19	0.88	0.30	0.30	0.10
2005	705047	8004	1.14	0.17	0.70	0.27	0.21	0.04

For the investigation all subjects, levels and grades were combined. As entries increased significantly over the five years, we used percentage of entries as a criterion. The question is now: At what value would the percentage be acceptable and can we say that the quality of marking was at an acceptable level? That magical figure will indicate that marking was up to tandard—we just have to determine the figure and agree on it. One of the aspects that this research shows is to perhaps extend it further to include a much bigger database, such as at a national level.

Graph 1: Trends in Remarking



Graph 2: Trends in Increase and Decrease of Raw Marks



In order to see where remarking had a significant effect, the remark analysis was extended to measure changes with remarking in 2% intervals, based on the raw marks. As can be seen in the graph, the tendency is towards a normal distribution around the 2–4 % interval. This brings us closer to an ideal situation that could become our standard.

Changing trends in remarking does not happen automatically. There has to be deliberate intervention actions and strategies. A very valuable contribution to the overall success of the marking process was making the detailed remark results available to Internal Moderators and Chief Markers. Not only could they determine how they actually fared, but could see in which intervals the problems were. During remarking sessions Chief Markers are now noting areas with major errors, not only those made by markers, but also sections of the work or memorandum that lends itself to inconsistent marking. Markers can now also be alerted to bad practices during the training session and put corrective measures in place to counter these. At the memo discussion the examiners can now not only engage in discussions about agreeing on the answers to accept, but also contribute towards meaningful to the marking process. Because we included analysis per subject paper and per subject, an added bonus from the year 2004 onward is that we are now able to pinpoint discrepancies to a specific paper.

It was found to be a useful tool for target setting to motivate markers. Subjects that set themselves formal targets after receiving the first analysis have made huge improvements. It was also used to recognize good performance and practices with subjects that did well.

Analysis of remarking results of Physical Science

The analysis was done for the 2004 (889 candidates) and 2005 (798 candidates) Physical Science Higher Grade answer scripts of the Senior Certificate Examination. The average

difference of the original and remark scores for the individual questions where determined. All scripts where there was a difference of two or more marks in any one question, or a difference of four or more on the total, were looked at individually to determine the nature of origin of the difference. The variety of different scenarios was just too diverse to use in pattern determination—mainly due to the unique nature of partially correct individual answers.

There were however some aspects that could be identified as 'culprits':

- Errors during carry-through of marks of incorrect/partially correct answers. This 'error' seems to be equally advantageous and disadvantageous to the candidates.
- Remark markers do not see a 'fresh' script and there is a tendency to repeat the error of the initial marker. A significant number of cases were seen where the remark marker was 'influenced' by the original marker to allow partially correct alternatives not provided for in the memo. This tendency was also observed in work that was moderated. The question may well be asked: Is the remark marker 'moderating' the original marking, or looking for places to add or deduct marks, or marking the script afresh, strictly according to the memorandum? The nature of the marker as educator will probably steer the marker towards advantaging the candidate.
- A complex marking memorandum makes marking difficult—too many alternative methods for with different partial marks are given in an attempt to have a comprehensive memo and this may well confuse the marker—or is it perhaps that the subject knowledge of markers is so poor that we need to give step by step instructions for everything?
- Surprisingly there were errors in marking Multiple Choice Questions—these were however related to candidates not following instructions, e.g. two crosses and one not cancelled clearly, cancelling a question and changing the numbers of subsequent questions on the answer sheet, etc.
- A major factor, usually picked up on during rechecking, is that candidates answer questions in bits and pieces. While it is generally picked up during marking, carry-through of answers and also contradictions and double dipping often get missed during initial marking. This aspect seems to contribute significantly to changes in marks.
- Deteriorating handwriting and letters and digits that have to be interpreted rather than read is a major culprit and is equally applicable to markers and candidates.

Questions, where the average difference between original score and remarked score was in the order of ≥ 0.1, were studied in detail. It is interesting to note that these relate to specific sections.

Table 2: Average difference between original and remarked scores in sections Pr. 1

Physical Science HG Paper 1 – Physics		
Section	**2004**	**2005**
1. MCQ's	0.02	0.01
2. Vectors	0.16	0.13
3. Forces	0.28	0.21
4. Motion	0.07	- 0.01

5. Laws of Motion	0.07	- 0.14
6. Energy	0.09	0.09
7. Momentum	0.07	0.12
8. Electrostatics	0.01	0.24
9. Current and Resistance	0.03	0.03
Total	0.81	0.68

Sections 2, 3, 5: These sections had the common scenario that there are various alternatives that can be used, requiring a very complex memorandum, containing carry-through of incorrect/partially correct answers. It generally required higher order skills from candidate.

Section 7: Trust a candidate to really mess up something when a concept has to be expressed in words or symbols. Language skills of both candidate and marker are tested and both groups get it wrong.

Section 8: Once again candidates cannot express concepts in words. Carry through into more than one alternative method caused problems. An interesting note on the memo stated: "weight accepted this year, but not in future"—examiners themselves are not consistent in requirements and a marker may remember a comment such as this two years from now and mark it wrong even if it is acceptable in the particular memorandum of a different section.

Table 3: Average difference between original and remarked scores in sections Pr. 2

Physical Science HG Pr. 2 – Chemistry		
Section	**2004**	**2005**
1. MCQ's	0.01	0.02
2. Intermolecular Forces and Gasses	- 0.30	0.24
3. Inorganic Chemistry	0.00	0.06
4. Reaction Rate	0.00	0.04
5. Chemical Equilibrium	0.39	0.04
6. Acid-Base Reactions	0.04	- 0.13
7. Electrochemistry	- 0.06	- 0.02
8. Organic Chemistry	- 0.03	- 0.02
Total	0.00	0.24

Section 2: Language ability of both markers and candidates are questionable. Candidates use 'correct' and 'incorrect' words in an specific context and markers do not always pick up on finer points – haste in marking may play a role where markers are only reading the words and not the context or way it is used. Formulae are often manipulated in an unusual way and then seem incorrect at first glance.

Section 3: Writing balanced equations and getting it partially correct confuses markers, despite instructions—mark allocation varies from year to year and question to question and markers start 'mixing instructions'

Section 6: More than one method can be used—candidates 'combine' methods, doing both correct and incorrect work in their total answer. This gives a problem if the candidate doesn't get it correct and has to be awarded partial marks.

Conclusion

All indications are positive that remarking can be used as a tool to quality-assure marking processes. Not only does it take away the subjectivity that goes with monitoring and observation, but is in fact one of the few quantitative measuring methods in education that are not subject to the ability of the specific group of candidates.

Further research into the statistical patterns in remarking, with a broad database, will lead to accurate criteria with which to evaluate the marking process. Research using the existing national papers as base will enhance the position on a national level. The move in RSA towards national question papers will make it possible to then establish a national standard against which marking can be evaluated. As remarking is a process where marking errors can be corrected, candidates should not subject it to an attempt to "gain" from the system. We believe it should be affordable, but candidates should feel the full effect of the challenge and accept the outcome, even if it results in a lower score. This should help to so that only real cases of a difference of opinion on quality of initial marking are evaluated against the norm and further be linked to qualifying criteria.

References

- Addis Ababa University, 2006. Registration Information. www.aau.edu.et/Applyonline/Reginfo/Acadmicslist.php
- City University London, 2006. University Information and Procedures
- www.city.ac.uk/edc/programmes/apractice/proregulations
- DEFS (2006) Department of Education and Skills. Access to Scripts. www.dfes.gov.uk/qualifications/
- Demopoulos, K. (2005) 'Rise in number of exam papers regarded'. In *The Guardian*. http://education.guardian.co.uk
- DOE **(**2005). Regulations for the conduct, administration and management of assessment for the Senior Certificate. Government Gazette (484) No. 28156. 21 October 2005. Pretoria. Government printers.
- GDE (2006). Circular 1 of 2006. Re-checking, Re-marking and the Viewing of the Answer Scripts of the Subjects Written in the Oct/Nov 2005 Senior Certificate Examinations. www.education.gpg.gov.za/Legislation/circulars/circulars.htm
- IEB (2006). General Notes: Senior Certificate Examination. www.ieb.co.za/Formal/FETC_remarks_2005.htm
- Open University of Hong Kong (2006) Course Results Query and Appeal. www.ouhk.edu.hk/
- Purpletrain (2006) PurpleTrain.com. Examination Queries. www.purpletrain.com/studserv/faq_exam.asp?ccode=0136
- Teachernet (2006) Appeals procedures for GCSE and A level examinations www.teachernet.gov.uk/educationoverview/uksystem/examinationsandqualifications/appeals/

- Technikon North Gauteng (2006) Remarking of scripts. www.tng.ac.za/operations/studadexam.html
- Tomlinson, M. (2002) Report on the Outcomes of Review of A-level Grading. www.dfes.gov.uk/
- University of Papua New Guinea. (2006) Rules and Procedures for All Undergraduate Programs. www.upng.ac.pg/uni_rules.htm)
- University of Warwick (2006) Frequently Asked Questions. www2.warwick.ac.uk/fac/soc/celte/welt/faqs/
- UP (2006) University of Pretoria. Access to Marked Examination Papers & Remarking of Examinations.

Looking for a needle in a haystack

Establishing 'standards' at the end of the GET phase (Grade 9)

Anne Oberholzer, IEB

Abstract

The Grade 12 year and the Senior Certificate examination dominate education in South Africa, encouraging schools to focus their efforts at the FET phase. Consideration of issues at earlier phases, particularly GET, is tempered by the role it plays in preparing learners for the FET. In that process, some key educational issues that pertain at the junior secondary school level are clouded by the pressures of the FET phase and perhaps, are not debated in their own right as GET issues, but rather are given attention only in so far as they impact on the FET.

Furthermore the GET has been an area of policy instability which has dominated and detracted from the educational debates that are relevant at this stage of secondary schooling. The pre-occupation of the South African society with Senior Certificate results at Grade 12 has under-mined the relative importance of educational issues lower down the system. There is a strong argument that suggests that national pass rates (and the associated quality of education!) at Grade 12 would improve if attention were given to some key educational issues at an earlier stage in the life of a learner.

This paper will consider what it means to establish a 'standard' at this level and the implications of establishing a 'national' standard at the end of Grade 9. It will include an exploration of necessary considerations in determining a 'standard' such as purpose, environment, systems and processes. It will offer some options for consideration in establishing a 'national' standard at this level. The IEB conducted some research into the GET phase and the explorations in this paper will reflect the responses from teachers to some of the issues.

The GET band and related assessment and qualification

Education policy documents immediately prior to and since 1994, the advent of democracy in South Africa, have spoken of the introduction of a General Education and Training

Certificate at the end of Grade 9. As matters stand, the first issue of the GETC for Schools will be in 2008.

In the SAQA GETC policy document the purpose of the GETC qualification is described as follows: The primary purpose of the GETC is to equip learners with the values, knowledge and skills that will enable or enhance meaningful participation in society, contribute towards developing sustainable communities, provide a basis for learning in further education and training, and establish a firm foundation for the assumption of a productive and responsible role in the workplace. (SAQA, pg 16)

Inherent in this statement of purpose are three key aspects: citizenship development, preparation for the world of work and laying the foundation for further learning. There is a perception that one qualification cannot do justice to all three aspects and hence one finds that there have been a number of GETCs registered on the NQF, each one with more specific purposes that emphasise a different aspect. Page 10 of the SAQA policy document provides for multiple GETC: Each GETC will provide access to various learning pathways, both vertical and horizontal, in terms of the purpose of the qualification. The scope of access provided by each GETC will be determined by the qualification itself. For example, the General Education and Training Certificate: Basic Technical Practice: Energy has the following purpose: The qualification enables learners to further their studies within the Energy sub-disciplines at NQF level 2 and above. Through the electives component of the qualification learners are able to demonstrate vocational skills and values through which they are able to engage in life skills activities, small business development, health and environmental issues. This is a workplace-based qualification that provides a basis for further learning within the energy sector.

In the early discussions about the National Qualifications Framework and its role in human resource development, the view prevailed that, after 10 years of compulsory schooling, learners would leave the formal schooling system and engage in meaningful alternate education and training pathways of learning either in revitalised technical colleges or through learnerships. That would require a set of comparable qualifications at the end of the GET band, to enable portability and progression decisions to made on the basis of demonstrated competence. The truth of the matter is that the revitalisation of the technical colleges has not happened as swiftly as one may have wanted. As a result of that delay, the qualifications offered in the colleges have not had an opportunity to gain currency in the marketplace or in institutions of higher learning and hence do not appear to offer a viable option for young learners as yet. However, more significantly, parents are not convinced that their children are best served by leaving formal schooling and entering learnerships or the technical college sector. The slow implementation of learnerships for younger learners has also prevented any meaningful alternative to formal education be it in schooling or colleges. So while compulsory schooling may technically end after 10 years, the reality is that children remain at school until the end of Grade 12 wherever possible. The dreams of a more flexible and vibrant system of human resource development has yet to be realised in South Africa.

The uncertainty about the relevance and role of the GETC (Schools) has contributed to policy instability and engendered a scepticism within the education community as well as

learners and parents about the proposed qualification. This scepticism has been exacerbated by the fact that the purpose of the qualification in schools has not been clearly identified. It has not been registered on the NQF and furthermore, it could be argued that any potential value of the qualification has been undermined by poor implementation strategies in respect of the formal assessment at the Grade 9 level, over the past few years. The credibility of the system at this level may well have been damaged quite substantially, with many participants in the system, from teacher unions to officials, calling for a complete rethink on government policy at this level.

In the late 1990's it became clear that masses of learners were not going to be leaving the formal education system at this level. Furthermore it was clear that the system would not be able to sustain a full-scale examination at this level as well as at Grade 12. It was also clear that the Grade 12 examination would remain the key assessment point that society in general believed to be credible, for the foreseeable future. There were many political questions about what purpose would be served by large-scale assessment at Grade 9—what if there was a massive failure rate at Grade 9? For whatever reason, there was exploration of what alternative assessment could be implemented to avoid the problems posed by a full external examination, such as the Grade 12, but would still provide a useful benchmark at the Grade 9 level. The response is the current policy provision for assessment at Grade 9 which allows for school-based continuous assessment (CASS) in 9 learning areas (the language learning area accommodates home and 1st additional) which is moderated externally and which constitutes 75% of the final result as well as an externally set and internally marked Common Task of Assessment (CTA) that constitutes 25% of the final result.

The primary purpose of the CTA is to have a common task that is conducted in all schools with the same requirements and assessment criteria so that all learners are assessed against the same criteria i.e. it provides a benchmark for comparison. In theory, by comparing the results achieved by learners in the CTA against their results in the school-based assessment, a school will be able to determine whether learners are being assessed at an appropriate level.

The difficulty in acquiring reliable data from schools across the country has meant that limited research has been done on the efficacy of the CTA as an assessment instrument at this level. The Gauteng Department of Education has done a report on the assessment of learners in Grade 9 in 2005 and as part of its research into a Core Skills Test at the Grade 9 level, the IEB is also looking at the data it has for the CTAs, the school-based assessment and the results in the Core Skills Test. This is a separate project from the one to be discussed in this paper i.e. the Grade 9 CTA Review Project, a more qualitative study.

The intention of the Grade 9 CTA Review Project was to get a sense of the perceptions of teachers and school managers about the Grade 9 IEB CTAs and the surrounding educational and policy issues. Questionnaires were distributed to all schools registered for the GET at the IEB i.e. 146 schools. The questionnaire dealt with generic questions about the CTAs as a whole as well as questions about the GET Curriculum and its introduction in schools. There were also questions that were specific to the CTAs in the different learning areas. A total of 762 questionnaires were returned to the IEB and analysed.

This paper will discuss aspects of the responses received in an effort to see whether we can find a needle in a haystack i.e. what do we mean by 'standards' at the Grade 9 level. The full report is available on the IEB website (http://www.ieb.co.za).

The question of standards

Teachers were offered an opportunity to make general comments on a variety of issues in the questionnaire. Among the comments made by teachers, there was one interesting observation as follows: the CTAs are really unnecessary as an instrument to establish a common standard because teachers know the standard required at Grade 9.

However, a review of responses within the questionnaire suggests that this may not be the case. It is clear that each teacher has a perception of the required standard, however, what is not so clear is how comparable these perceived standards are. Assessment literature is full of discussions on what actually is a standard—there is little agreement on this, but substantial agreement that 'standards' is a term which is more loosely used than any other in education (Gipps, 1990). In the questionnaire there were a number of opportunities for school managers and teachers to comment on standards and it is interesting to see what emerged.

In one question, school managers were asked to rate the learning areas in terms of consistency of demand by indicating whether, in relation to the standard their school sets at Grade 9, the standard of the CTA was easier than is required, of the correct standard or more difficult than is required. The following table is a summary of their responses to the question:

	% rating it as easier than is required	% rating it as being of the correct standard	% rating it as more difficult than is required	% who offered no response
Arts and Culture	**19**	**29**	**5**	**47**
Economic and Manage-ment Sciences (EMS)	**20**	**28**	**3**	**37**
Human and Social Sciences (HSS)	**15**	**24**	**16**	**46**
Language, Literacy and Communication: Home	**13**	**34**	**15**	**38**
Language, Literacy and Communication: 1st Additional	**15**	**37**	**3**	**45**
Life Orientation (LO)	**11**	**36**	**7**	**46**
Mathematics	**7**	**27**	**25**	**41**
Natural Sciences	**15**	**40**	**4**	**41**
Technology	**11**	**30**	**10**	**49**

From the table, one can deduce that in the perceptions of teachers, it is possible the level of demand in the CTA in Arts and Culture, Economic Management Sciences and 1st additional language may have been too easy and in Mathematics the level of demand may have been too difficult. It should be noted that all CTAs were evaluated by Umalusi external moderators who considered them as being of an appropriate level for Grade 9.

However an interesting observation about the rest of the responses is that in 4 of the learning areas, there are as many school managers who think that the CTA in a learning area was too easy as there are school managers who think that it was too difficult! For instance, in Human and Social Sciences 15% thought it was easier than is required and 16% thought it was more difficult than required; In Home Language 13% thought it was easier and 15% thought it was more difficult, and in Technology 11% felt it was easier and 10% felt it was more difficult.

Of interest too, is the rather large number of respondents who did not respond to the question. While one may be tempted to deduce that this means that a substantial group of school managers do not know whether standards are being met by the CTA, the key reason however is likely to be that the grade co-ordinator is not a subject expert in all learning areas and hence is not comfortable with passing a comment on the standards in learning areas with which he/she is not familiar.

It is worth looking at the teachers' responses in each of the learning areas to questions about the standard of the 2005 CTA. In response to the statement "the CTA was set at the correct standard to be expected from Grade 9 learners", the following responses were indicated by teachers of the specific learning area:

Learning area	% that agree	% that disagree	% that don't know / no response
Arts and Culture	**70**	**30**	**0**
Economic and Management Sciences (EMS)	**53**	**43**	**4**
Human and Social Sciences (HSS)	**48**	**50**	**0**
Language, Literacy and Communication: Home	**53**	**41**	**6**
Language, Literacy and Communication: 1st Additional	**64**	**32**	**3**
Life Orientation (LO)	**69**	**25**	**5**
Mathematics	**47**	**50**	**2**
Natural Sciences	**72**	**24**	**4**
Technology	**77**	**20**	**3**
All teachers	**61**	**35**	**3**

The range of responses that indicate that the CTA is not of the correct standard is significant across all subjects with the greatest disagreement being in Human and Social Sciences and Mathematics (50%) and the lowest discrepancy being in Technology (20%).

The most important deduction from this table, as pertains to the question of standards, is that there is no learning area in which 80% or more of teachers agree on whether the CTA was set at the correct standard—the highest percentage is 77% and the lowest is 47%.

In response to the statement "This year's CTA was more difficult than previous CTAs", the following responses were indicated:

Learning area	% that agree	% that disagree	% that don't know / no response
Arts and Culture	**36**	**45**	**19**
Economic and Management Sciences (EMS)	**30**	**43**	**27**
Human and Social Sciences (HSS)	**51**	**27**	**22**
Language, Literacy and Communication: Home	**58**	**22**	**20**
Language, Literacy and Communication: 1st Additional	**21**	**60**	**19**
Life Orientation (LO)	**39**	**33**	**28**
Mathematics	**44**	**26**	**30**
Natural Sciences	**11**	**64**	**25**
Technology	**33**	**41**	**26**
All teachers	**36**	**40**	**24**

This table illustrates comparisons within the learning area across two years. Again it is valuable to note that teachers do not think alike in comparing standards. Across all learning areas the pattern is the same: some teachers believe that the CTA was more difficult than in previous years while a significant percentage of teachers, in the same learning area, disagree.

Quite interesting is the large number who did not respond to the question. One can speculate about what that means. An interesting thought is that it could be an indicator of that there is less stability in the teaching force that one might see as desirable at this level - they cannot respond because they were not involved with Grade 9 CTAs in the previous year!

In response to the statement "This year's CTA was easier than previous CTAs", the following responses were indicated:

Learning area	% that agree	% that disagree	% that don't know / no response
Arts and Culture	**16**	**61**	**22**
Economic and Management Sciences (EMS)	**19**	**51**	**30**
Human and Social Sciences (HSS)	**12**	**65**	**23**
Language, Literacy and Communication: Home	**5**	**70**	**25**
Language, Literacy and Communication: 1st Additional	**31**	**46**	**23**
Life Orientation (LO)	**13**	**57**	**30**
Mathematics	**6**	**63**	**32**
Natural Sciences	**35**	**39**	**27**
Technology	**11**	**59**	**30**
All teachers	**17**	**57**	**26**

Even in a subject like Mathematics where there are a number of different indicators that suggest that the 2005 CTA was difficult, there are still teachers who do not think so—6%

believe that the 2005 CTA was easier than previous years and 47% believe it was set at the correct standard to be expected from Grade 9!

The question that began this discussion was: do Grade 9 teachers know what the standard is that should be expected from Grade 9 learners? From the responses in the questionnaires, it is not possible to say that there is a common standard that all teachers understand. Even within learning areas, teachers do not agree about the standard of the 2005 CTAs; they do not agree that the CTA in their learning area is set at the required standard expected of Grade 9, nor do they agree whether it is easier or more difficult than in previous years.

The obvious follow-on question is this: is it possible to establish a generally expected standard at Grade 9? Personally I am not sure, given the fact that despite its long history, debate continuously rages about what the standard at Grade 12 is and whether it is acceptable for tertiary education. Perhaps an exploration of what contributes to establishing a standard would be helpful.

How does one establish a standard?

The starting point would seem to be an in-depth consideration of the purpose is for setting a standard. The Senior Certificate examination doubles 'as a school leaving certificate and a university entrance qualification' (NEPI, 1992, pg 25). Is its purpose to measure learners' competence at a specific level, in which case the system is determining minimum requirements, or is its purpose for selection and reward purposes, in which case the system is looking for the best fit to a specific set of criteria, or is its purpose simply to identify those with potential in which case we are using part performance to predict future success. Some argue that one assessment process cannot do all three, while others disagree. Whatever the academics may say, there is general acceptance of the outcomes of the assessment process at Grade 12 and how they are used in society.

Official documentation at Grade 9 is not direct about the specific purpose of a GETC—neither with respect to the issue of minimum competence or with respect to selection for continuation of learning or placement elsewhere in the system!! In the 1998 survey carried out in the KZN Department of Education, 76% of teachers offered an opinion about what purpose is served by the issuing of a GETC. Many of those responses reflected the early rhetoric about the GETC i.e. it would enable those who get a GETC to go to work; it would force the education system to concentrate on preparing students for the world-of-work, developing skills other than traditional academic skills. There were others that indicated that it would re-focus public attention elsewhere in the system and not only at Grade 12, and ensure comparable standards at this level and hence uplift the standard of education generally. Some cited its purpose as the establishment of a comparable minimum competence level across different sectors, while others indicated that it would provide a mechanism for selection of learners for further education or career pathing.

Responses from teachers in the IEB survey ranged from a view that there is no purpose for a GETC to the opinion that the GETC provides a valuable benchmark in the system, depending on the quality of the external assessment intervention. Interestingly, both groups identified that the GETC would encourage alternate forms of assessment, with the IEB

survey highlighting teacher development in respect of outcomes-based assessment! Given the agitation and rumblings of dissatisfaction about the current state of the GET, it was surprising to note that 56% of managers indicated that their school would like to provide successful Grade 9 learners with a GETC certificate.

The most direct statement about the GETC qualification for the school sector and its purpose is in the Overview document for the Revised National Curriculum Statement Grades R-9 (Schools), in which it says (DoE, pg 3) that the "General Education and Training Certificate, aligned to the Revised National Curriculum Statement, is due to come into force in 2008 and this certification is set to mark the end of compulsory schooling". However the document does not provide any clear direction as to what this means—it simply indicates that "the assessment and certification of this accumulative learning take place at the end of Grade 9 when learners who meet the specified requirements will be awarded a General Education and Training Certificate." (DoE, pg 19).

A consideration of the CTAs may provide some insight. While 69% of all respondents indicated that the CTA was a valuable educational activity and 59% of respondents indicated that the CTA was a valuable tool for teacher development, there were a variety of responses as to what the purpose of the CTAs is:

- 80% of teachers in the survey agree that the key purpose of the CTA is to provide a common task against which schools can compare the standards demanded by their own SBA against a common standard.
- 78% of respondents are of the opinion that the purpose of the CTA is to provide teachers with examples of how they might assess the achievement of outcomes.
- 73% believe that its purpose is to broaden the types of assessment tasks used in schools.
- 35% believe that the purpose of CTAs is to replace examinations with an easier alternative and ensure learners 'pass' Grade 9.
- 15% indicated that the purpose of CTAs is to provide learners with something to do to keep them busy until the end of the year!

From this set of responses, it is clear that the teachers are not unanimous about why CTAs have been introduced.

We should not be surprised at this—if the purpose of the qualification is not clear, then it is likely that the purpose of the assessment instrument will be unclear. If this is the case, then too one must assume that there is little hope of agreement of what should be assessed and how. Furthermore if neither teachers nor learners know why the assessment is taking place, there is a huge question mark over the validity and reliability of the outcomes of the assessment process. The motivation of teachers and learners must also be an issue.

If we look at what should be assessed, the policy documents indicate that the new GET curriculum is the basis for the GETC assessment, both in respect of SBA and the CTAs. While the explicit introduction of learning outcomes into the teaching and learning process was praised, the 2005 Curriculum Review process identified the under-specification of content one of the problems with implementation. The IEB survey also shows that teachers themselves may be confused about what they think of the new Curriculum! 79% of

respondents indicated that the new GET curriculum requires more from learners than simply mastery of subject knowledge. 59% of the respondents are of the opinion that the new GET curriculum makes significant intellectual demands on learners. However 41% indicated that the new GET curriculum has contributed to a decrease in the knowledge and skills our learners have at the end of Grade 9. Quite high percentages were evident in the respondents from Mathematics (51%) and Arts and Culture (58%).

The apparent contradictions in the analysis of these responses beg further interrogation. For example, it would be an interesting study to investigate how the new GET curriculum has contributed to a decrease in the knowledge and skills learners have at the end of Grade 9, if at the same time, it requires more than mastery of subject knowledge and makes significant intellectual demands on learners. There are a variety of possible explanations for this seeming contradiction—the reduced specification of the content knowledge required at Grade 9 has meant that some teachers in lower classes have avoided teaching what could be deemed as critical content and hence the learners are not up to the standard required. Again, the perception that the clustering of subjects into learning areas has reduced the development of discipline-specific skills may be responsible for the belief that this has contributed to a decrease in the knowledge and skills learners have at the end of Grade 9.

Umalusi has consistently emphasised the importance of more detailed specification of the knowledge and skills required when trying to quality assure an external assessment process. They have maintained that if you don't have a clear idea of what is to be assessed i.e. the specification of the syllabus, there is no way of establishing a standard. One must be cautious and tread carefully over these eggshells—one of the criticisms of criterion-based assessment is the over-specification of what constitutes acceptable responses, to the extent that the desired outcomes of learning are subsumed in descriptions of details. Note the words of Edmond Holmes, written in England in 1911:

> The state is prescribing a syllabus which was to be followed, in all subjects of instruction, by all schools in the country, without regard to local or personal considerations, was guilty of one capital offence. It did all his thinking for the teacher …. He should gradually deliver himself, mind and soul, into the hands of officials of the Department….. The teacher who has been deprived by his superiors of freedom, initiative and responsibility, cannot carry out his instructions except by depriving his pupils of the same vital qualities. (quoted in Gipps, 1990, pg 104).

The balance between useful structuring of a framework and over-specification is possibly more critical to education than any other profession.

In setting standards, a critical consideration is established practice which depends on communities of practice and trust. These are developed over a period of time, where professionals confer and engage in debate in an effort to deepen their understanding and exploration of contentious issues—standards is a contentious issue! South Africa's divided history has not been helpful as there has until the last ten years, been minimal engagement between teachers. In 1998, 49% of teachers in the survey carried out in the KZN department believed that teachers will try to advantage their candidates by purposefully giving candidates marks that are higher than they actually deserve. A further 11% believed it will happen

depending on the professionalism and integrity of the individual teacher (KZN, 1998). Even now there are those who argue that serious engagement between teacher in different areas of operation is virtually non-existent. Without engagement it is unlikely that trust can be developed.

One of the current practices in South Africa, to try and build communities of trust is the system of clustering. This is what teachers in the IEB survey had to say:

- 69% of school managers believe that CTAs provide an externally set assessment instrument that assists in judging internal standards
- 56% of managers indicated that the CTAs provide a valuable benchmark for schools within each of the learning areas.

However

- 82% of teachers and school managers indicated that the Gr 9 cluster meetings are important to establish a common standard across schools;
- 66% of responding teachers indicated that the Grade 9 cluster meetings have been valuable for them as teachers while 83% indicated that the Grade 9 cluster meetings have the potential to improve the performance of teachers and learners. There were a substantial number of individual comments from teachers reflecting on the very positive experiences of their cluster groups – sharing of experiences and tasks, discussion, and learning. There were some comments that reflected either negative experiences in clusters or a complete disagreement with the philosophy of cluster groups;
- 72% indicated that the clustering process ensures an ongoing check on what teachers are doing. While this may seem to be unnecessary and should be the responsibility of the school managers within a school, the re-assurance from a cluster meeting that what is happening in other schools is comparable to what is happening in one's own school, is an important aspect of building up communities of trust – an important factor in building a common understanding of what a 'standard' is.

There are numerous international examples of attempts to build communities of practice e.g. exemplars, banks of items, publication of learner responses in external assessment to illustrate standards at different levels. These are all useful, but I would contend that it is only through direct engagement that teachers develop communities of practice and hence, trust. The IEB experience of the introduction of portfolios seems to verify this. Ros Janisch is presenting a paper on this very issue at this conference.

Conclusion

If we are going to establish a credible standard at the GET level firstly we need to:

- Acknowledge that the pre-occupation of the South African society with Senior Certificate results at Grade 12 has under-mined the relative importance of educational issues lower down the system;
- Recognise that there is a strong argument that suggests that national pass rates (and the associated quality of education!) at Grade 12 would improve if attention were given to some key educational issues at an earlier stage in the life of a learner;

- Recognise that the GET phase has not received the attention it deserves if it is to make a real difference in building a quality education system.

Then we need as an education community, to clarify the purpose of any assessment process at Grade 9—is it to identify weaknesses in the system at an earlier stage? Is it to improve and establish accountability earlier in the system? Is it to mark the end of compulsory schooling and hence indicate minimum competence? Is it for selection purposes?

Once that is clear, we need to determine what should be assessed and then the best assessment process to address the purpose or purposes that have been agreed. These are hard questions that require hard decisions. The ball is in the court of the practitioners who need to actively contribute to these debates.

Finally if we intend to establish any credible standards at the Grade 9 level we need to develop communities of practice among teachers, who through their consensual understanding of the requirements establish standards that are trusted by all.

References

- Department of Education (2002) Revised National Curriculum Statement Grades R-9 (Schools) Policy Overview
- Gauteng Department of Education (2006) Draft report on the assessment of learners in Grade 9 in 2005
- Gipps, C. (1990) *Assessment: a teachers' guide to the issues*. Hodder and Stoughton
- IEB (2006) Grade 9 CTA Review Project Report
- Lubisi, C. and Murphy R. (2002) 'Assessment in South African Schools'. In *Assessment in Education*, 9(2) July
- KwaZulu-Natal Department of Education (1998) Analysis of questionnaires received from 18 schools in the KwaZulu-Natal Department of Education Year-mark project (unpublished internal document)
- National Education Policy Initiative (1992)
- South African Qualifications Authority (2001) The General Education and Training Certificate (GETC) Policy Document.

"A tail of many dogs"

Assessment and the scramble for standards

Sandile Ndaba, Umalusi

Abstract

Educational assessment increasingly focuses on uniform educational standards and accountability for students and teachers in order to meet common educational standards. Examinations are used for a variety of purposes: to help learners, parents and prospective employers to make educational and career choices, to provide a basis for selecting individuals for specific sorts of further or higher education, or for certain jobs, and for qualifying individuals to undertake specific vocational or professional activities as well as to exclude those that do not qualify. It is for this reason that those involved in examinations often refer to examinations, pejoratively, as 'the tail that wags the dog'. However, this 'tail' that is examinations has many dogs (the various contending purposes to which examinations are put) and it wags them in accordance with the various agendas of educational technocrats.

This paper discusses the issue of standards and the use of examinations in the 'scramble for standards'. It begins by discussing the term 'standards' in general and then proceeds to consider standards in the South African education system. It argues that examinations, while they serve extremely important purposes of licensure, certification and accountability, cannot, however, ensure quality education. The author argues that too much is invested in the ability of examinations to drive standards. In the course of this, important quality levers are forgotten.

Introduction

This paper explores the pervasive and hegemonic dominance of assessment in reform efforts to improve the standards and quality of education. Current educational reforms in many countries around the world call for the establishment of clear performance standards for schools, and results of large-scale assessments are increasingly reported in terms of the percentages of learners ready to meet minimum standards. A crucial question is whether standards-based reporting is adequately understood by non-technical audiences—such as the press, policy makers, and the public—that are among the most important consumers of assessment results.

A stream of reports and pronouncements has fuelled the popular perception that the South African education system is in crisis. Real and imagined declines over time in performance in

various forms of assessment; such as the Senior Certificate Examination, the systemic evaluation tests, cross national comparisons on tests like the South Eastern and Eastern Consortium for Measuring Educational Quality (SACMEQ), and comparisons to benchmarks such as TIMSS and MLA have been publicized as evidence that our educational woes continue unabated.

The use of test scores to index educational success or failure is rarely ever questioned. Low scores are bad news; and high scores are good news. In the rhetoric of education reform, it often sounds as if improving the education system is synonymous with improving test scores. In such a climate, the logic of high stakes testing is compelling. The logic is so seemingly simple: Test learners and see what they can do and hold schools accountable if they fail to make the grade. Rather than micro-manage schools, policy makers can dictate that content standards and performance standards be created to codify expected educational attainments.

The question posed in this paper is whether this obsession about testing as a means to improve educational quality and standards works.

The issue of standards

The term 'standards' has become an important rallying point for educational technocrats. It has become emblematic of reform efforts aimed at improving quality and standards in education. The concept is, however, notorious for its elasticity, elusiveness and malleability. The use of the term in the public rhetoric often masks its real complexity. It can have several meanings:

- A standard can relate to a criterion score on a test. Both the methods of scoring and the decision about setting the criterion level depend on somebody's judgement of what is satisfactory or what constitutes a high or, worse still, a world class standard.
- A high standard could refer to the height of the hurdle or to the proportion of students who can jump it and there is an obvious trade-off between the two. A common restriction of meaning is to equate high standards with test performance.
- A national standard might be a minimum of performance that all should attain, or a scale of performance levels on which each individual will be placed (Black, 1998: Linn, 2003; John, Hoover and James, 2004).

Conceptions of standards can, therefore, be based on content standards which describe in specific detail what content should be covered, how and at what level of complexity. Sometimes standards are perceived as measures of outcomes where specifications about the competencies to be displayed by learners and their levels of demonstration are specified. Standards are also conceived as the quality of practice inside the classroom; often referred to as 'opportunity to learn'. Finally, standards may be conceived in terms of the quality of inputs or resources devoted to the system in terms of schools, teachers and instruction. Yet when we talk about standards, we talk as if we all understand the same thing.

The drama on standards plays itself out in the public psyche in deeply symbolic ways on a variety of stages, enacted by different characters and it demonstrates varied texts and

subtexts that are nonetheless all embedded in the meta-text of quality education. Although punted for the ostensible reason of improving the quality of education, standard-based assessments serve mainly to hold those involved in education to account.

The tail and its many dogs

The main aim of assessment is to provide evidence of the quality of performance in a given outcome. This evidence is based on judging a carefully selected sample of that performance. This does not seem to raise too many problems. However, the moment we begin to think about the reasons for assessing and why we assess in one way rather than the other and the purposes to which the outcomes of assessment are used, many problems arise.

Assessment can be used for a variety of purposes: to help learners, parents and prospective employers to make educational and career choices, to provide a basis for selecting individuals for specific sorts of further or higher education, or for certain jobs, and for qualifying individuals to undertake specific vocational or professional activities as well as to exclude those that do not qualify.

All the aforementioned, in my perjorative and rather irreverent inversion of the common reference to assessment as the ***tail that wags the dog***, are the 'dogs' that the tail that is assessment wags.

The main complexity that arises out of the purposes to which assessment is used is when it is used to hold those involved in education to account. The internal uses of assessment for diagnostic purposes, or its formative uses to improve the learning of those assessed are less problematic than when it's used for accountability. Assessment for accountability is a 'high stakes' affair.

I will now consider how assessment has been used as a pawn in the game of standards.

Assessment and the scramble for standards

To add insult to injury, assessment is frequently used to draw conclusions on standards and the quality of education. Assessment becomes the main driver of standards and quality in education: the tail that wags the dog, as the common parlance among assessment professionals refers to it. It is used interchangeably in both formative and summative ways to make judgements on standards. Assessment scores are often interpreted through making a complicated set of comparisons:

- Comparisons over time
- Comparisons between and among examining boards
- Comparisons across different subjects
- Comparisons across and among schools
- Comparisons across and among districts and provinces
- Comparisons across and among racial groups
- International surveys
- National surveys.

This desire to use assessment to drive the quest for standards and quality education is a worldwide phenomenon manifested in the surveys, conducted internationally, regionally and nationally to measure the quality of educational attainment.

In the late 1950s the International Association for the Evaluation of Educational Assessment (IEA) was formed. It initiated what would be major studies aiming to measure the quality of cognitive achievement at various levels of education in several countries and to identify the main causes of differences in achievement of educational outcomes. Twelve countries joined its first mathematics study. By 2000, fifty countries were participating in surveys covering mathematics and science, now called the Trends in International Mathematics and Science Study (TIMSS), science and reading, known as Progress in International Reading, Literacy Study (PIRLS).

The IEA surveys precipitated a scramble for standards. Strongly influenced by the IEA experience, several other such studies, usually of regional focus, have since been established. These include the Programme for International Student Assessment (PISA), set up by the OECD in 1998 and now covering fifty-nine mainly first and second world countries; the Southern Eastern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) which, since its first survey in Zimbabwe in 1991 has expanded to fifteen African countries; the Latin American Laboratory for the Assessment of Quality Education (LLECE), which began in 1997 and covers sixteen countries; the UNESCO Monitoring Learning Achievement (MLA) project and the survey in French-speaking Africa known as PASEC. (EFA Global Monitoring Report, 2005: 44-46)

Not to be outdone, South Africa has joined the fray to participate in some of these surveys typifying the scramble for standards: TIMSS, SACMEQ and MLA. As if this was not enough, two national surveys called systemic evaluation at grades 3 and 6 have so far been conducted and more are in the pipeline.

The question that is still begging is: **Has this madness about testing helped to improve the quality of education?** Not really. That would be my answer. The immense potential that assessment has to drive instruction is undisputed. However, many educationalists are less confident of its potential to improve educational quality (Chisholm: 2004, Muller: 2004, Kanjee: 2005, O'Neil: 2005, Wolf: 2005, Lolwana: 2006). Tests of cognitive achievement are incomplete proxies for the quality of education.

The results of these surveys are, however, very instructive on factors that influence educational quality and standards. I will now turn to some of these factors.

Factors influencing standards and quality of educational attainment

Each study made great efforts to identify the most salient and significant factors that influence standards and quality education (EFA Global Monitoring Report: 2005). All the studies referred to above show that nearly all education systems, pupils' home background

proved to be crucial. Learners from higher socio-economic backgrounds where parents had higher levels of education and were thus materially well off tended to perform better than their counterparts from poorer homes. In Africa and Latin America, the urban-rural divide exacerbated the problem.

In many developing countries, the material resources in schools are inadequate. This comes out very clearly in the SACMEQ studies where the average child was in a school with 8.7 of the twenty two desirable school resource items, with the disparities ranging from 4.3 items in Malawi to 16.7 in the Seychelles, this complicated by wide variations in urban-rural variations within countries. Even in those countries where a relative degree of equity in the distribution of material resources had been achieved, the teachers in urban schools tended to be better qualified than their rural colleagues.

The gender of primary school teachers was found to have a profound influence on performance; particularly of girl learners. The SACMEQ studies showed wide discrepancies by country. For all countries, 53% of pupils surveyed were taught by female teachers, on average, but the share ranged from 17% in Uganda to 99% in the Seychelles. Learners taught by females scored, on average, three tenths of a standard deviation higher than pupils with male teachers.

Other items prominent in the African studies were behavioural problems of pupils and teachers: late arrival, absenteeism and learner dropout were all correlates of poor performance. In the PISA studies where socio-economic school climate, teacher morale and commitment, school autonomy, teacher-pupil relations and disciplinary regime had some compensatory influence towards greater equity. I Latin American countries covered in LLECE studies, pupil socio-economic background and classroom climate appeared to be the most important predictors of achievement.

Forty years of research conducted through the IEA studies reveal some major findings relating to the quality of educational attainment. Three of these findings are particularly instructive for policies aimed at improving educational quality. The first is that the distribution of cognitive abilities in the population has a significant impact on average achievement levels. The greater the overall proportion of children enrolled, the lower the average achievement levels tend to be. The achievement levels of particular cohorts of ability, however, are not affected—the cognitive achievement levels of the most able cohort remain unchanged by expansion. The second is that time spent actually working on particular subjects, either in school or as homework, affects performance, especially in mathematics, science and languages. The third is that although socio-economic status is influential in determining achievement in all contexts, textbook availability and school resources appear to be capable of countering socio-economic disadvantage, particularly in low income settings.

Conclusion

Reform efforts aimed at raising the standard of education have relied too heavily on using summative assessment as a lever. However, these efforts have not been as successful as it was hoped they would be because they ignored the formative purposes of assessment. A

small, but growing group of researchers are now turning their attention to assessment embedded in the context of the classroom. They argue that more emphasis should be laid on formative assessment as this allows for feedback to the learner and thus enhances the opportunity for learners to improve their performance. The cumulative effect of this will be a steady improvement of the learners' overall performance which would in turn improve the quality and standards in education. While standards-based assessment may tell us how many learners can hurdle the bar we set for them, it contributes little to the strategies for improving teaching and learning in the classroom.

References

- Black, P (1998) Testing: Friend or Foe? Theory and Practice of Assessment and Testing. London: Falmer Press
- Chisholm, L. (2004) 'Public Debate Lacks Quality Information and Analysis'. In *HSRC Review*, 2 (1) March 2004
- John, J, et al. (2004) 'Differentiating Standards-Based Education for Students with Diverse Needs'. In *Remedial and Special Education*, 25 (2) March/April 2004, pp 74-78
- Linn, R.L. (2003) Performance Standards: Utility for Different Uses of Assessments. In *Education Policy Analysis Archives*, 11(31). Retrieved 19 May 2006 from http://epaa.asu.edu/epaav11n31/
- Lolwana, P. (2004) 'What is the Standard of the Senior Certificate Examination?: Umalui's Report in Context'. A Paper Delivered at the HSRC Matric Colloquium on the 3rd of November 2004.
- Muller, J. (2004) 'Assessment, Qualifications and the National Qualifications Framework in South African Schooling'. In Chisholm, L (ed). *Changing Class: Education and Social Change in Post Apartheid South Africa*. Cape Town: HSRC Press
- O'Neil, O. (2005) 'Assessment, Public Accountability and Trust'. A Paper read at the First Cambridge Assessment Conference, 18 October 2005
- Wolf, A. (2005) 'What can and should we be Measuring?' A Paper read at the First Cambridge Assessment Conference, 18 October 2005

Standards—a couple of steps back

Mark Potterton, Umalusi

Abstract

Throughout the world there is obsession with 'educational standards'; and a lot of effort goes into the development of standards in the belief that these improve the quality of education. This paper briefly explores what standards are, and what their limitations might be. An argument is developed that standards do little to change the complex processes of teaching and learning, particularly in contexts of poverty and deprivation. The author concludes that besides the obvious need for basic inputs in developing country schools, there is a need for teachers to engage students, develop their imaginations and inspire them to do great things.

What are standards? Standards are about measurement. They are sometimes used to refer to what students are able to do at what stage of their schooling: what words they should be able to spell, what facts they should be able to recall or what mathematical problems they should be able to solve. These standards are then incorporated into tasks and tests that can be used to measure the extent to which students can meet them. These standards are further used as a lever in standards-based educational reform to improve the quality of education.[1]

Introduction

The standards debate has dominated the education reform movement in the United States (Darling-Hammond 2004), and the rhetoric of this movement has been that students will not be able to meet the needs of changing economies if they do not have more challenging work in school (and there is truth in this, particularly if there are limited jobs available). Some

[1] Standards are generally written by experts and displace locally developed curriculum. These 'standards' are generic and don't always understand the contexts of rural and other marginalised communities.

believe that once clear standards are spelt out on what students should be able to do and know then it is easier to mobilise other resources to support student learning.[2] Darling-Hammond (2004:1081) believes, and I agree with her, that standards and accountably can't be separated from school organisation, teaching, assessment, staff development and resourcing issues.[3] Changes need to be made to the entire structure of the system to improve student learning. Standards and authentic assessments can be used as indicators of 'progress' in the system but not as sanctions for schools who under perform.

But can we really talk about standards in examination systems on a continent where many schools don't have textbooks or other books? Or where children don't have enough space to move in their classrooms? Or where many children come to school hungry or ill? In this paper I asked that we take a step back and establish what we need to put right in our schools before we worry too much about standards in examinations.

Imagine yourself for a moment on a mountain dirt road. It's 7:30 in the morning and laughing children walk past you on their way to school. The children carry their books in plastic bags in one hand and in the other they have bottles of water, as there is no running water at the school. You can see the school in the distance, and as you get closer you see girls sweeping out the classroom. After some prayers outside in the cold the children take their places in the crowded classroom. Three children, and sometimes four, squash into a desk made for two. The teacher greets the children and asks them to copy the notes from the chalkboard. The chalkboard is well-used and it is difficult to read from it. The teacher, a young woman who has recently qualified, isn't employed by the government. She teaches in her private capacity and is paid a small wage by the school. The students copy carefully into their dust covered notebooks. The teacher can't move around and check what they are doing because there is no space. In the background the voice of the other teacher who shares the classroom space can be heard explaining the other 'class'…

Later that day the young teacher recounts the history of Lesotho. She speaks of the great Moshoeshoe and how he united a divided people. She describes how he defended his people from Zulu raids and preserved their independence against Boer and British interlopers. The children sit and listen. The children parrot "King Moshoeshoe rallied the people together in 1820". Some of the children you passed on the way to school are dosing off. Lerato is gazing out of the window, and Mamello is punching Ramakeele. James is hungry, as he hasn't eaten yet. And the teacher fails to engage the students; and misses the opportunity to talk about democracy and community and other important things.

This vignette illustrates a very real experience of a school in southern Africa, and points to significant features in schooling in the developing world. Most obviously, it shows the deep

---

[2] Lefkowits et al (2004) note that standards are very low on the United States public's lists of priority concerns about schools. Issues like discipline, drug use, overcrowding and the lack of funding are of higher concern than standards or teacher quality.

[3] Elmore (2004: 209) argues that the standards and accountability movement in the United States is being transformed into a testing and accountability movement because some States don't have the money and human resources to select, administer and monitor tests at all grade levels. He notes that this is the surest way to guarantee that the test becomes the content.

inequalities in the world, and reminds us that poverty is a painful reality for many people and impacts significantly on schooling.

In South Africa we have blazed ahead introducing a new curriculum complete with assessment standards. We have published glossy boxed supporting documents, carried out very short teacher development courses and hoped that things will improve. But I don't believe that enough attention has been paid to the context of reform or to what happens in classrooms. I will now take a couple of steps back and briefly consider the impact of poverty and HIV in schools; I will then review some classroom-based research and consider more closely what teachers are doing.

Poverty and HIV and schools

Poverty and HIV impact on the life of schools in southern Africa and can prove a stumbling block for any standards-based reform.

In an article about schools in disadvantaged contexts in Britain, Ruth Lupton (Lupton 2004) showed how these contexts impacted on the organisation and processes of schools and that these differed from one area to another. Lupton writes about deprivation in Britain, but from my experience in the field (as well as from the literature for example Levin & Lockheed, 1993), these findings are equally true in southern Africa. One of the areas of impact was the wide range of abilities within each school. This placed additional demands on teachers trained to teach in mainstream schools and to teach students in English. Teachers tended to make use of telling rather than writing, and worksheets and copying exercises were used more regularly in these classes.

Another area of impact in these schools was material poverty. The lack of resources limited the range of extra-curricular activities. For example, enrichment trips had to be heavily subsidised and planned in advance so that parents could budget to cover the cost. The assignment of homework had to be considered carefully as so few, if any, of the students could be assumed to have learning resources like reference books or computers at home. Students did not always arrive at lessons with basic equipment like pens or rulers, which meant that time was spent during lessons giving out or taking in stationery.

The third and possibly the most distinctive impact according to Lupton (2004) was that the emotional environment of all these schools was highly charged: "The number of students who were anxious, traumatised, unhappy, jealous, angry or vulnerable was reported to be much greater than in schools where parents were materially well off, less stressed themselves and more able to secure a stable and comfortable environment for their children" (Lupton 2004: 9). In some schools students were disruptive, finding it difficult to concentrate and to accept rules, and were sometimes aggressive towards other students and teachers.

The emotional needs of students had a wide impact on schools. Students tended to share their needs with their teachers, and teachers spoke about 'mothering', 'caring' and 'social work'. This meant that teachers had to develop strategies to deal with these emotional needs in the classroom. For example, more time was spent on dealing with individual needs rather than on teaching the whole class. Teachers had to develop clear routines and introduce new

challenges carefully as many students were threatened by change. This meant that dealing with students' emotional and behavioural problems often took precedence over teaching and learning.
Teaching in these contexts was draining and demanded more on a personal level than just teaching a subject. Teachers often had to deal with drama, conflict and tears, and found it hard to remain detached.

Attendance at the schools was poor. Schools with extremely disaffected students tended to have low overall attendance rates. Parent involvement in meetings and other school activities was low. Other day-to-day issues for teachers included having to cajole students to complete homework, and to return books and equipment to class.

Lupton (2004) noted that all of these issues together resulted in an 'unpredictable working environment'. Something could happen at any time and lessons could not be relied upon to go according to plan. Student mobility also meant that it wasn't clear how many students would be at school on any given day.

HIV infection too, impacts on many facets of the school-going child's life (Utley, Potterton & Potterton, 2005). There are social, cognitive, emotional, physical and behavioural implications impacts on each student. There are also many factors that influence how a student who has HIV will be affected, including the home environment, access to health care, poverty, and family support and structure. No one picture fits all children and each student's relative strengths and weaknesses must be considered.

Students who have HIV are often malnourished, and they may find it hard to eat because of sores in their mouths. They may also have difficulty digesting food and absorbing nutrients because of all the gastrointestinal tract infections they have. This means that they are often very small and thin for their age. HIV-infected students are often weak and lethargic because they are sick so often and are malnourished. These students sometimes tire quickly at school and do not always have the energy to participate in physical activities.[4]

Students who are infected with HIV are faced with many social and emotional challenges. They often do not have the support that they need in order to lead carefree childhoods. Many of these students become depressed. They withdraw socially and show loss of initiative and interest in participating in activities. They may show symptoms of fatigue, lethargy and inattentiveness and may be less likely to express their emotions.

Behavioural problems are not uncommon in students who have HIV. These children are often stigmatised and bullied at school and may respond with high levels of aggression. As

4 HIV can get directly into the brain where it slowly destroys certain important cells. This destruction of brain cells means that it becomes harder for students to learn new skills and, if it continues, they may even start to lose skills which they have already acquired. This includes motor, speech and cognitive skills. These changes can be emotionally difficult for children as well. Anti-retroviral drugs have been shown to slow down or even halt this process. If children do not get treatment they will become progressively more disabled and over a period of time these children may become blind, deaf and unable to walk or talk.

children reach adolescence they become more aware of the impact of the disease on their futures and sexual behaviour.

HIV does not affect a child in isolation but is a disease of the family. Students are often forced to assume care-giving roles for ill parents or younger siblings as well as dealing with their own illness and trying to cope and fit in at school. Poverty and poor access to adequate health care are additional challenges many families face.

Into the classroom

Let's stop here and look a little closer at what might be happening in Lerato's and Mamello's classroom. Over the past six years I was involved in efforts to monitor the impact of a literacy and numeracy teaching in rural schools in Limpopo, Eastern Cape and KwaZulu-Natal. We tested students and observed teachers both before and after they participated in training, mentoring and resource enrichment opportunities. The post-implementation study made use of classroom observations[5] and a questionnaire about management and governance of the school. The teachers we observed worked in challenging contexts:

- Some of the classes were large and conditions were cramped.
- There was a lack of teacher reference and other reference materials.
- There was a shortage of storage space in some schools.

We identified a number of positive aspects in the schools and these included:

- well-kept schools
- good relationships between students and teachers
- punctual teaching staff who spent the required time at school
- teachers who were open to new ideas and being mentored
- good classroom management skills.

Most of the teachers arrived on time for lessons and started teaching immediately. A number of teachers had made an effort to display posters and use other teaching aids. The observations showed that teachers still dominated classroom activities. Whole-class teaching was the main approach used by most teachers and in some instances teachers used interactive whole-class teaching. More teachers were using assessment in their lessons than in 2000, and many were beginning to record a range of assessment tasks

We also noted that there were teachers absent in most of the schools that we visited, and a number of students were also absent. Some principals reported that the impact of HIV was beginning to be felt on attendance levels.

Lesson planning

There was evidence across all of the schools that lesson planning was taking place, but that this planning was only adequate. Lesson plans should really be developed with the primary purpose of planning what learning will happen, and not just for the sake of having them.

[5] A team of seven of us observed 222 lessons in 19 schools.

Lesson plans also needed to pay attention to the types of activities that the students would engage in. Lesson plans also needed to show how the lesson would be introduced and how the lesson would be concluded. Lesson plans ought to have provided a connection between the knowledge and skills being taught and how they will be assessed.

Whole-class teaching

Whole-class teaching was the most commonly used form of teaching. Whole class teaching can be useful when introducing new content and providing instructions and to set expectations. It was also useful to introduce the same skills to the whole class.

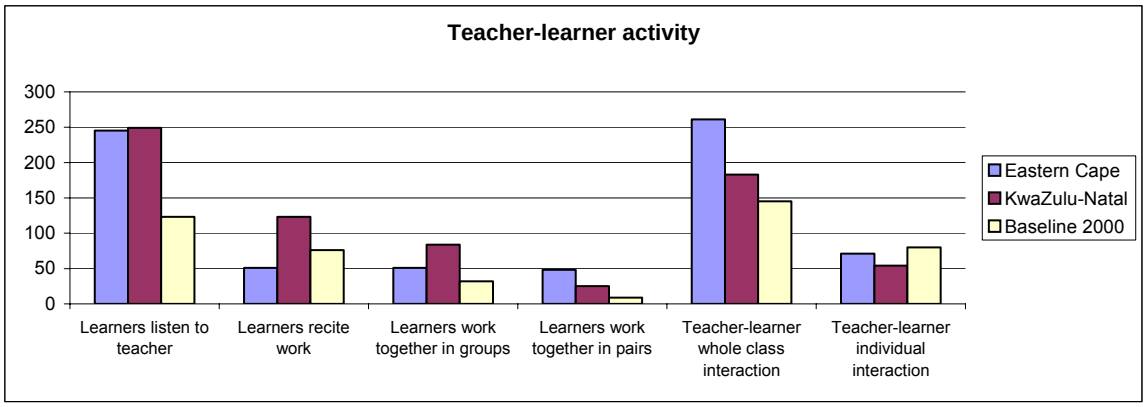


Figure 1: Teacher-learner activity

The large classes and limited space also tended to force teachers to mainly use whole class teaching. However, too much teacher-talk was counter-productive as students found it difficult to concentrate for extended periods of time, and different teaching approaches would have created variety and encouraged students.

Teachers needed to mindful to ensure that whole class teaching strategies were used to promote learning. Teachers should speak, say for 10 minutes, the students work for 15 minutes, and teacher speak again for 10 minutes. Teachers needed to create opportunities for students to talk and share their ideas, and students should feel freer to ask questions when they do not understand what is going on. Teachers would probably have found that students write more effectively after they had first presented their ideas verbally.

Classroom management

In general, teachers were proficient at managing large groups of students. A number of classes had classroom rules displayed that spelt out clear behaviour expectations. Teachers were able to use their voices well to control unruly behaviour and a few of them used humour to good effect. Unfortunately, schools still did not have very good systems in place to manage discipline, and a lot more attention can be paid to establishing reward systems that affirm and celebrate student achievements, rather than only relying on corporal punishment.

Many of the teachers still had students copying notes from the chalkboard. This approach was reinforced by the shortage of textbooks and the lack of duplicating equipment. By

introducing a variety of teaching approaches teachers could have engaged students better and made learning interesting.

Promoting critical thinking

One of the main tasks of teachers should be to promote thinking skills at schools. Teachers could set up special areas that get students thinking, for example displays related to current news issues. Schools in other parts of South Africa have set up problem corners with a problem and students are then encouraged to pin up solutions. The purpose of exercises like this would be to encourage students to ask questions and suggest solutions.

Teachers generally expected immediate answers to their questions. If they didn't get quick enough answers then they usually answered the question themselves. The students were not active participants in the schools, and teaching for thinking could have made these schools more interesting and more challenging places for the students. Better still, teaching children to think requires no expensive teaching aids.

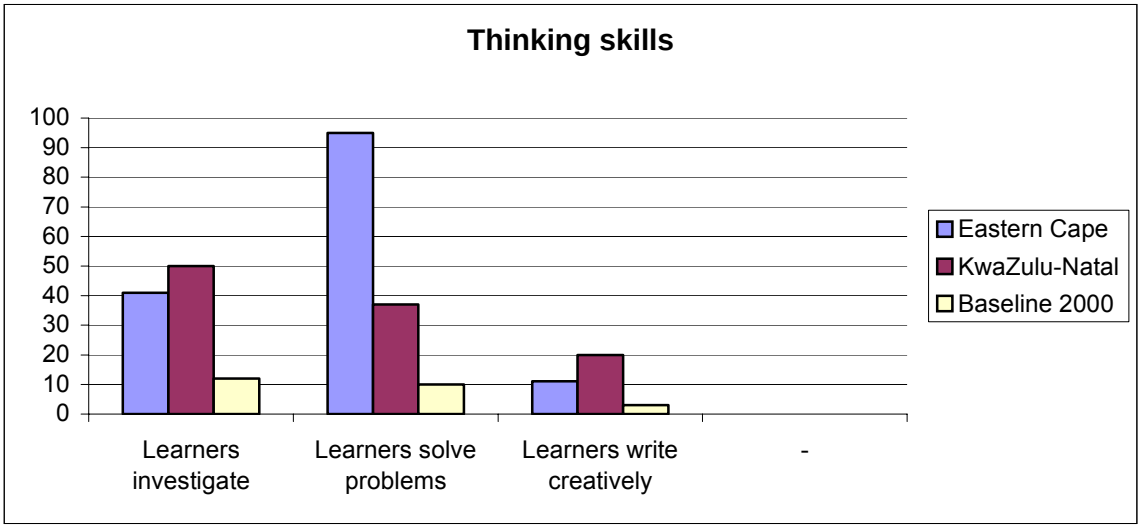


Figure 2: Thinking skills.

Assessment

Assessment was poorly managed in most of the schools. Assessment in outcomes-based learning should provide an indication of individual student achievement, which wasn't there. Assessment needed to ensure that students integrate and apply both knowledge and skills, and assessment needed to be used to allow students to make their own judgements about their own performance and progress. Assessment also needed to inform teaching practice. In other words, teachers needed to use assessment to identify gaps and to change their teaching strategies. In short, teachers need to increase the range and types of assessments and develop profiles of individual student performance and development.

The findings in our post implementation study still showed that there was too little written work being completed, that the pace and sequencing of learning was poor and that little attention was being paid to progression. There was still too much rote learning and very little cognitive challenge. This is not good news. In Goodlad's study (1984) of more than a

thousand United States classrooms he too found that students listen, read short sections of textbooks, and respond briefly to end-of-chapter questions or worksheet questions. Students seldom analyse their mistakes or get an opportunity to try again until they have 'learnt' the material. There were few opportunities to analyse information or assess different points of view.

Linda Darling-Hammond (1997) reporting on research in the United States argues that students who learn small bits of information and 'parrot back' this information in the style in which it was learnt have little ability to use the information in new situations or to connect this information across subjects. She says that students who continually learn by rote are eventually not able to reason things through or to estimate or judge whether an answer is plausible. In fact she agues "they may actually understand less rather than more as they proceed through the curriculum" (Darling-Hammond 1997: 54). She goes on to argue that students engaged in hands-on learning opportunities like projects, discussions and research which encourage high-order thinking are better able to remember and apply what they have learnt than students who have learnt by rote:

> Teaching aimed at understanding and applying ideas once though appropriate only for selected high-achieving students – has proved more effective than rote teaching for students across a wide spectrum of initial achievement levels, family income levels, and cultural and linguistic backgrounds (Darling-Hammond 1997: 55).

Unfortunately teachers who are tied to learning programmes or textbooks often don't stop to explore the interesting ideas through debates, discussion, newspapers, contemporary politics or additional reading. Because teachers are keen to cover as much as possible they neglect to develop real understanding. Information is dealt with in a superficial way. In Africa large classes and a shortage of learning materials and library resources are often given as the reason for this superficial approach to learning.

Towards a conclusion

Standards and assessment, and the accountability that the combinations of these bring need to be questioned. We must be sensible. We have to maintain our focus on students like Lerato, Mamello and James and ask what schooling does for them. Does a new curriculum with all of its associated paraphernalia marginalize poorer children further? Or does it provide the necessary knowledge and skills to open up new opportunities for students like Lerato and Mamello?

Curriculum standards must mean more than just a list of bulleted sentences for the young teacher in the overcrowded classroom. Standards need to go hand in hand with school building programmes[6], nutrition programmes, learning materials development and meaningful staff development which looks at things like:

[6] Elmore (2004: 89) argues for reciprocity in accountability, for example: "For every increment in performance I require of you, I have a responsibility to provide you with the additional capacity to produce that performance." He continues to argue that: "Test-based accountability without substantial investments in capacity – internal accountability and instructional improvement in schools – is unlikely to elicit better performance from low-performing students and schools…it is likely to aggravate the existing inequalities between low-performing and high-performing schools and students (Elmore 2004: 207).

- Planning and implementing simple lesson exemplars with clear outcomes.
- Teaching through a variety and balance of teaching styles.
- Organising and structuring the learning environment.
- Promoting higher levels of cognitive demand.
- Developing the active participation of students (engaging students, accessing prior knowledge etc).
- Ensuring that lessons have a clear application activity that can be assessed[7].

Acknowledgements

The assistance of the people and groups who participated in the 2004 post implementation study must be acknowledged. Special thanks also need to be extended to the principals who so willingly assisted in the study.

References

- Darling-Hammond (1994) 'Standards, Accountability, and School Reform'. In *Teachers' College Record*. 106 (6), 1047-1085.
- Darling-Hammond (1997) The Right to Learn: A Blueprint for Creating Schools that Work. San Francisco: Jossey-Bass.
- Goodlad, J.I. (1984) *A place called school*. New York: McGraw-Hill.
- Hargis, C.H. (2006) 'Setting Standards: An Exercise in Futility'. In *Phi Delta Kappan*, January 2006, 393-395.
- Fuller, B. (1991) Growing-Up Modern: The Western State Builds Third-World Schools. New York: Routledge.
- Lefkowits, L. & Arens, S.A. (2004) The National Dialogue on Standards-based Education: Final Report. Washington: Regional Educational Laboratory.
- Levin, H.M. & Lockheed, M.E. (1993) *Effective Schools in Developing Countries*. London: The Falmer Press.
- Lupton, R. (2004) Schools in Disadvantaged Areas: Recognising Context and Raising Quality. London: London School of Economics.
- Potterton, M., Utley, E., Goolab, C., Shoko, B., Twyford, E. Dawjee, J. & Bodenstein, M. (2004) Report on the post-implementation classroom study in the Eastern Cape and KwaZulu-Natal. Johannesburg: CIE Education Services.
- Potterton, M., Utley, E. & Potterton, J. (2005) Mid-term Evaluation of the Window of Hope: HIV and AIDS Prevention and Care Programme. Johannesburg: CIE Education Services.

Appendix 1: Professional development: the consensus view (adapted after Elmore 2004: 96)

- Focuses on a well-articulated mission or purpose anchored in student learning of core disciplines and skills.
- Derives from analysis of student learning of specific content in a specific setting.

[7] This list can easily be expanded and should definitely include improving the salaries and benefits of teachers (see Appendix 1).

- Focuses on specific issues of curriculum and pedagogy.
- Derives from research and exemplary practice.
- Connects with specific issues of instruction and student learning of academic disciplines and skills in the context of actual classrooms.
- Embodies a clearly articulated model of learning.
- Ensures collaborative practice within schools.
- Sustains a focus on continuous improvement over time.
- Provides models of effective practice.
- Uses assessment and evaluation.
- Actively monitors student learning.
- Provides feedback on teacher learning and practice.

Determining educational standards in Malawi

Duncan MC Nyirenda, Mzuzu University

Abstract

Although the Malawi Education System does not have a document entitled Education Standards, the standards are implicitly spelt out in different Government papers on education.

The Policy Investment Framework (PIF) stresses quality education. Quality is determined by specific prescribed benchmarks of success within the education system. Such benchmarks describe what level of achievement should be attained in order to achieve quality education; these benchmarks can, therefore, be equated to standards.

Basic education in Malawi has been described as "to equip students with basic knowledge skills to enable them to function as competent and productive citizens in society. There are therefore, these elements of knowledge and skills obtained from various disciplines that enable learners to be productive citizens in society. The purpose of Secondary Education on the other hand is to provide the academic basis for gainful employment in the formal, private and public sectors. It will also prepare students for further education. The paper will attempt to illustrate what implicit standards have been developed to achieve these goals

The paper outlines such skills as literacy and numeracy that pupils should attain. With the HIV scourge, life skills education has become very prominent in the system so as to educate the pupils about the dangers and prevention of HIV and AIDS. The paper further discusses the formulations of the committees which design and develop curriculum for different levels of education. These committees formulate the level of attainment. These level of attainment prescribe standards for that levels.

The Quality Assurance body is the Ministry of Education Inspectorate Department commonly known in Malawi as Educational Methods Advisory Services. The paper outlines and discusses some roles of the Department as regards to quality assurance and standards.

Introduction

Malawi does not have a written document called Education Standards as it is in other African countries such as Uganda. The desired education standards are spelt out in various official documents of the Ministry of Education. Such documents include syllabuses, the Policy Investment Framework (PIF) and Vision 2020.

To begin with, what is meant by Education Standards? 'Standards' is a highly problematic concept that can bear many different meanings or interpretations. Sometimes 'standards' refer to the academic attainment of students: in the new syllabus for primary schools in Malawi, there is what we call 'Assessment Standards'. This refers to the level of attainment that is expected of primary school pupils. This is the concept of standards that the general public in Malawi means when they talk of, for instance, 'standards have gone down'. Their concern is how well students are performing. If the term 'standards' focuses on students' attainment, it then means knowledge, skills and values related to a chosen subject (Kelly AV 1990). While it is common knowledge that when there is the question of standards of education, the focus becomes on students' performance and the process of teaching and learning. Other inputs into the education system which uplifts the standards in education are ignored. The Malawi Policy Investment Framework (PIF) has outlined the inputs that constitute what standards in education entail. These are discussed below:

The inputs that the policy addresses are:

Physical infrastructure

The space on which the school is located should be enough to accommodate school blocks. A library, laboratory (secondary schools), pit latrines, etc, and a field for games, such as football, net ball, basket ball, etc. The standard area is 2.5 hectares. In some cases, this could be larger because teachers' houses are also included in the physical infrastructures. A head of institution's house should be within the compound. The critical input however is the teaching area. The structures must provide a conducive environment for teaching and learning. The Ministry of Education stipulates the standard classroom size. This standard classroom should accommodate a teacher/pupil ratio of 1:60. There are moves within the Ministry to reduce the classroom size at primary school level to 1:45 by the year 2015.

There have been a lot of self help projects on school construction. The government provides model schools in each district that local communities would emulate in school construction. These schools still exist but it has been extremely difficult for the communities to emulate because there have been inadequate financial resources and the escalating prices for construction materials.

There has been massive support for the provision of standard infrastructure from several donors. DFID has constructed several standard school structures across the country. They are currently constructing school blocks and a teacher's house in Ntchisi District. Ntchisi has about 130 primary schools. GTZ under the Basic Education Project in Zomba has helped in renovating school blocks in both rural and urban areas. European Union, Christian Council of Malawi, USAID are some of the organisations that have played a big role in helping Malawi to meet the development of infrastructure in the school system.

Furniture

The minimum furniture standard is 1.2 pupils in primary schools. It is therefore expected that every pupil sits on a chair. This is an area of great concern. Most schools, particularly primary schools have very low desk pupil ratio. The table below exemplifies the situation as it exists in primary schools:

Table 1: National Pupil Desk Ration

Standard	Enrolment	Desk Ratio
One (1)	804 312	10 116
Two (2)	547 457	10 858
Three (3)	472 337	15 006
Four (4)	353 974	18 205
Five (5)	302 410	28 052
Six (6)	214 155	32 397
Seven (7)	168 605	38 052
Eight (8)	153 722	46 231

(Source: Education statistics 2000)

Although there is a prescribed standard of pupils' furniture, there is still a critical shortage of desks. The situation is a little better in secondary schools where there is at least a ratio of 1:2; one desk to 2 pupils. No student sits on the floor in all Government and Government aided secondary schools. The situation is, however, different in private secondary schools. In the majority of these schools, three or four students share one desk and in some instances students actually stand or sit on the floor.

Personnel Requirements

Qualified teachers

In order to attain minimum standards of education at both primary and secondary level, the policy stipulates that there must be a supply of adequate qualified teachers. The standard teacher/pupil ratio in primary school is 1:60 and 1:45 in secondary schools. In some districts, teacher/pupil ratio in primary schools is as high as 1:134. The shortage is estimated at 15 000 qualified primary school teachers. This is an indication that there is a great shortage of qualified teachers.

Training

Teacher training for primary school teachers is done in five Teacher Training Colleges. However, these colleges do not produce enough teachers to man the educational system. It is estimated that a total of 1 500 teachers graduate each year. This is far from being adequate considering the high attrition rate of 12% per year. The training programme would never catch up with the national requirement at the current rate due to 1). Insufficient space in colleges and 2) Financial constraints. The Government, however, is considering introducing a distance training for primary school teachers. This would double the number of primary school teacher graduates per year. At secondary school level the opening of Mzuzu University has considerably increased the output rate of secondary school teachers. University of Malawi, Domasi College and Mzuzu University graduate a total of about 500 secondary school teachers in a year. This number is inadequate to satisfy secondary school requirements.

Teaching and Learning materials

One of the determining factors of the Standard of Education is the adequate provision of Teaching and Learning materials. The standard of textbook provision according to Education Statistics (2005) at both primary and secondary school is at least 1:1. However, procurement of such materials has been problematic due to financial constraints. The table below depicts what is the reality of the situation on the ground. Grade one has deliberately been highlighted because this is where the highest enrolment in primary school is registered.

Table2: Pupil Book ratio

Subject	Std.1	Std.2	Std.3	Std.4	Std.5	Std.6	Std.7	Std.8
English	1.1	1.2	1.3	1.3	1.2	1.2	1.1	1.1
Chichewa	1.2	1.2	1.3	1.3	1.3	1.3	1.3	1.2
Maths	1.1	1.1	1.2	1.2	1.2	1.1	1.1	1.1
G/Studies	1.2	1.2	1.2	1.2	-	-	-	-
Life Skills	1.4	1.4	1.2	1.2	1.3	1.5	1.1	1.7
Science/ Health	-	-	-	-	1.2	1.2	1.2	1.1

(Source: Education Statistics 2005)

Curriculum Delivery and instructional process

Standards of education are set at the Curriculum Design and Development level. What pupils should learn and how, is determined during the curriculum design and development process. The depth and breadth of each subject at various levels are determined by curriculum panels. At both primary and secondary levels, curriculum panels draw up the scope and sequence charts and teaching syllabuses for various subjects at each level. The teaching syllabus indicates what level of attainment is expected.

Teaching syllabus stipulates the content, how the content should be taught and period allocated for each level. This is important in order that there is some uniformity in the teaching process. In a centralised curriculum, it is critical to set some minimum standards for teachers to follow, although evaluation theory in recent years has shown the evaluation process within education to be highly complex and sophisticated a matter, they have among other things, revealed the need for evaluation to be part of the process of curriculum change, and development from every outset: (Kelly 1990). The stress has been the value of formative evaluation in contrast to summative evaluation. Curriculum design and Curriculum process go through the formative evaluation for the sake of maintaining standards. This process continues during curriculum implementation where the inspectorate department takes on the role of maintenance of standards. However, there seems to be a widely shared concern about standards of learning in schools in Malawi. After the various inputs including financial provision, it appears there is no real agreement about what standards of the output of the school system at all levels should be. Indeed there have been expressions of dissatisfaction with academic attainment standards.

It appears a general concern is about the standard of learning and students' achievement in our schools. The system needs to produce certain standards of graduates. Within this production model of schooling there are two prerequisites according to Becher and Maclure (1978). The first is the agreement about which of all the things learnt in schools, are the most

critical learning. Becher suggests that it is important to ensure that the learning so chosen are critical so that particular attention should be paid to them during assessment. The second prerequisite is the agreement about what levels of attainment should be expected from pupils at different levels of learning. Becher in this thesis is focusing on the students' output model. We should, however, take cognisance of all other factors that constitute the standard of education.

Education and poverty reduction

Education in Malawi is a catalyst for poverty reduction. An educated populace can best participate in the economic, social, political and technological development. The purpose of education at each level should address these themes that relate to the poverty reduction. At primary school level the emphasis is literacy and numeracy. However, recently the government has come up with a broader purpose of education. According to the new purpose, the government has the responsibility to

> promote the overall development of all learners so that each learner becomes literate, numerate and has a basic understanding of science and technology, is a responsible, morally sound and productive citizen in a democratic society, who is also equipped with skills, values and attitudes to survive economically and socially and has a desire for a life-long learning. Perhaps it is important to stress that Primary education is terminal to about 80% of graduate.

It is therefore imperative that the emphasis at this basic level of education should be the acquisition of permanent literacy and numeracy. Consequently, the acquisition of permanent literacy and numeracy is the standard that the government would like to achieve. Literacy levels are perhaps the lowest in SADC at about 80%. There has been a general outcry that school leavers at both primary and secondary schools are not able to communicate because the literacy levels at primary and secondary schools are low. Despite the literacy and numeracy concerns, the government is focusing on the provision of essential skills and knowledge on a broader range of issues including critical thinking and analytical skills, civic and democratic values, modern technology skills, life skills and entrepreneurial skills. All these will add up in moulding a whole person into a productive citizen in the society. When these aims are achieved then the country would move towards the desired goal of standards in Education. An educated society would enhance the reduction of poverty.

Teachers and the delivery system

The basic standards that are required of teachers are:

- A scheme of work,
- A lesson plan,
- A record of learners progress,
- Attendance register.

The standard delivery of the lessons includes the participatory approach. Pupils should be masters of their own learning. The teacher should act as a facilitator.

Teachers' lesson plans show evidence of various techniques as appropriate for the lesson. These include objectives, groupwork discussion and evaluation of the lesson at the end. Some comments by the teacher on how the lesson(s) were conducted should be seen on the lesson plan or scheme of work. For effective learning each learner's progress is evaluated and records of the evaluation are in the pupils' record book. These are some of the basic fundamentals of teaching and learning process. However, the quality of teaching rests with the individual teacher.

Public confidence

Government is accountable for the provision of education to its citizens. Every individual particularly children have a right to education. Therefore, the government has to reassert the public interest in what is taught and to make sure that the results of the education match the **acceptable standard**. Since 1994 during the time free primary education was introduced, there has been a declining public confidence in the manner teachers have carried out their teaching responsibility. The decline in the confidence is partly as a result of the many untrained teachers that were employed because of the large enrolment of pupils due to free primary education. Coupled with the decline in the pass rate at National Examination, there has been a public outcry about the quality of education. The public notion of standard relate to pupils' performance. Normally pupils poor performance be it in examinations or in job market tend to reflect two major causes: poor quality of teachers and the inadequacy of teaching and learning materials.

Mention has been made concerning the teachers and the learning materials. These two and others that include financial resources affect the achievements of the set standard of education.

Maintenance of standards

The Ministry of Education has the overall responsibility of the maintenance of education. However, this responsibility is delegated to the Education Methods Advisory Service (EMAS). The department has two major responsibilities: Inspection and the provision of Advisory Services. The core business has been inspection. The vision of the department is "provide quality assurance feedback to all Educational Stakeholders in all Education Institutions". While the mission is to "establish, maintain and improve education standards" (EMAS briefings to the Parliamentary Committee on Education 2006).

The Inspectorate department has been decentralized in order to effectively perform the various inspectorate and advisory services. The country has been divided into six education divisions. The divisions are further decentralised into districts and finally there are 315 zones where there are zonal inspectorate and advisory officers. In each zone there is one officer who performs both Inspectorate and Advisory Services. This is a complicated setup and it needs proper and efficient co-ordination. The department looks forward to the time when all school and colleges would become self-evaluating institutions.

The major concerns in the inspectorate and advisory services include:

- student level achievement,

- relevance of learning—employment, social, cultural and political world,
- condition of learning,
- supply of teacher and learner facilities,
- provision of teachers' buildings, equipment and relevant curriculum. The department strongly believes that quality education is concerned with the process of teaching, learning, testing, managing and resourcing. All these are the ingredients of the standard of education. The department of Inspection and Advisory Services is within the limited financial resource endeavours to monitor the standard of education.

Malawi National Examination Board (MANEB)

MANEB is a National Examination Board. The Board's main mission is to conduct valid and reliable national examinations for certification, selection and placement purposes. The focus is on student achievement. The degree of students' achievement sometimes reflect implicitly the level of the education inputs. When the examination results are good, no questions are asked about what is happening in schools except where studies are conducted to emulate the good practice. But when the results are poor, there are questions such as: Are the teachers qualified? Are there adequate teaching and learning materials? Is there a good teacher/pupil ratio? What is the learning milieu like?

In a way MANEB implicitly monitors the examination standards and since they are present here, they will speak for themselves.

References

- Ministry of Education and Culture (2000) Policy and Investment Framework (PIF); Lilongwe, Malawi.
- Ministry of Education. Education Management Information System (EMIS)
- Education Statistics (2005)
- Nyirenda DMC (1985) A Framework for the Evaluation of Distance Instructional Materials at Malawi Correspondence College. A dissertation submitted in partial fulfillment of the degree of MA. (Ed.) by Examination and Dissertation. U. K. University of Southampton.
- Kelly, AV (1990) *The National Curriculum. A Critical Review*. London: Paul Chapman Publishing.
- Becher, Tony and Maclurel Stuart (1978) *Accountability in Education*. London: Nelson and Publishing Company Ltd.
- Ministry of Education and Sports (2001) Basic Requirements and Minimum Standards. Indications of Educational Institutions. Kampala, Uganda.
- Ministry of Education Science and Technology (2003) Malawi Primary School Draft Syllabus for Standard Six. National Curriculum and Assessment for Primary Education Framework.

‘Getting it right, keeping it right’

Ros Janisch, IEB

Abstract

Forcing learners to be creative and write essays under examination pressure is not considered best practice assessment. However, relying on the professional judgement of teachers depends on a community of trust. The high stakes of the Grade 12 Senior Certificate make teachers anxious and competitive. The inclusion of creative and transactional writing in the school based assessment component has brought into question how to maintain standards. Moderation procedures are in place and are working, but when the quality assurance of the results cannot include statistical moderation, it is difficult to convince the community that the results are valid and reliable.

This paper looks at the experience of the IEB over the last few years in running the writing portfolio. It looks at the methods used for standardisation, for teacher development, for the process of moderation and for feedback. It also outlines the reasons for the IEB’s wishing to continue with having creative and transactional writing assessed in the school based component rather than in an external examination.

Portfolios for Home Languages

> Educational assessment has … become divorced from learning, and the huge contribution that assessment can make to learning (ie formative assessment) has been largely lost. Furthermore, as a result of this separation, formal assessment has focused just on the outcomes of learning, and because of the limited amount of time that can be justified for assessments that do not contribute to learning, has assessed only a narrow part of those outcomes. The predictability of these assessments allows teachers and learners to focus on only what is assessed, and the high stakes attached to the results create an incentive to do so. This creates a vicious spiral in which only those aspects of learning that are easily measured are regarded as important, and even these narrow outcomes are not achieved as easily as they could be, or by as many learners, were assessment regarded as an integral part of teaching. Dylan Wiliam

When the IEB took over as an examining body from the Joint Matriculation Board in 1993, there was a unique opportunity to reflect on what should be examined in an external

examination for the Senior Certificate. The fact that systems of timed written tests and examinations narrow the curriculum was hardly news. Much of international literature on education at that time talked of the importance of coursework and the fact that some skills and knowledge were not suited to pen and paper examinations. Learners' portfolios were being included in many assessment packages and seemed to have great value. Learning was becoming the focus of assessment, and the previous stress on summative assessment, with its focus on external examinations, was being questioned as unfair.

External examinations have to be feasible and affordable, and therefore, by their nature rely on sampling as they cannot assess the whole curriculum. As Dylan Wiliam says:

> Giving so little time to the assessment means that we can assess only a limited proportion of what has been taught, and conducting the assessments in 'standardised conditions' means that teachers can easily guess which parts of the curriculum are going to be assessed. Because of the importance attached to these outcomes, teachers and students are pressured into focusing on only those aspects of the curriculum that will be assessed.

He goes on to say that in the UK, they started out with the aim of making the important measurable, and ended up making only the measurable important. I think all examination systems do this.

As Handy states when discussing the Macnamara fallacy:

> The first step is to measure whatever can be easily measured. This is OK as far as it goes. The second step is to disregard that which can't easily be measured or to give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily really isn't important. This is blindness. The fourth step is to say that what can't be easily measured really doesn't exist. This is suicide'. (Handy, 1994 p219)

The theory of Multiple Intelligences, the cynical view of summative assessment, and a desire to achieve more valid and authentic ways of assessing learners' competences have led to a recognition of the importance of assessment **for** learning, not merely **of** learning. Formative assessment became the buzzword of educators. I like the way formative assessment is defended in this statement by Peter Silcock:

> We shouldn't want it [a shift to formative assessment] because research shows how it improves learning (we don't need to be told that—it has to be true). We should want it because schools are places where learners should be learning more often than they are being selected, screened or tested in order to check up on their teachers. The latter are important; the former are why schools exist. —Peter Silcock

When discussions took place to design the examination requirements for the first IEB Senior Certificate examinations for English First Language (and later for all the languages examined at First Language level), discussions were held between the IEB and practicing teachers of languages at First Language level. The desire to involve some coursework was unanimous. As unanimous was the feeling that it was the assessment of writing that was least served by an external examination and that best suited coursework.

I am in full agreement with the words of Ben Brooks, retired headmaster of Redhill School in Johannesburg and English teacher for more than 40 years:

> Getting young people to enjoy reading and to write well is, in my opinion, the most important of the many challenges facing any first language educator. The two components are as two sides of one coin.

When reading about the need to encourage learners to write and to find their own voice in creative writing, I was struck by a memory of reading Charles Dickens' novel, '*Great Expectations*'. In this story of manipulation and revenge, the embittered old woman, Miss Haversham, seeks to repay the male gender for her having been jilted by her fiancé. She adopts a young girl and invites the main character of the novel, Pip, to come to her house. She intends to use Pip in her planned revenge. Pip comes for the first time to her gloomy, vast mansion and enters her large, dark room in which she sits in her wedding dress (years after the wedding date). He is overawed already, but is completely dumbstruck and incapable of responding when she orders him to 'play'!

What struck me was the similarity between this situation and the manner in which candidates for the Senior Certificate, in a large examination hall, under pressure of time and ambition to pass, are ordered to 'create!' and have to write an original essay. Not an ideal situation! Replacing an external examination with a portfolio of coursework in writing seemed an excellent idea. And so it has proved. The rationale for this approach in the teaching and examining of Writing in any first language is explored in many scholarly papers. I do not intend to summarise these here. I am merely going to list (briefly) **some** of its strengths:

- It is a process of continuous assessment.
- It encourages reflection, self evaluation and ongoing improvement.
- It includes editing, a crucial component of writing.
- It places the responsibility for assessment where it should be, in the hands of the educator who knows his or her learners well.
- It expects continual growth and innovation from educators.
- It facilitates synergy and sharing through user groups, cluster groups and peer moderation.
- It covers a great deal more than can be assessed in any examination.
- It is what writing is about in the real world, a whole range of registers and genres.

Teachers were immediately excited by the new approach and took the responsibility of supervising the coursework very seriously. The early years of the IEB First Language writing portfolios saw interesting debates: some teachers loved the challenge and emancipation of the writing portfolio; others missed the 'lift into excellence' that sometimes occurred in a learner's writing under examination pressure. Tasks for writing were designed with new enthusiasm, and, for the first time, the importance of the process of writing could be stressed. Portfolios of writing could showcase a learner's ability across a wide range of writing styles, demanding differing registers, formats and approaches. Not only was narrative, descriptive, discursive and argumentative writing demanded in the section for extended writing, but a wide range of transactional writing types were asked for in the

section for short pieces. A really exciting addition was the removal of one of the genres of literature from the external examination so that it could be explored and taught differently. Rigorous tasks on this fourth genre had to be included in the portfolio, but teachers were encouraged to avoid tasks that mirrored the essay and contextual type of the examination and seek intertextual investigations and cross-genre exercises. Some really interesting, original and exciting work has resulted here.

One of the first problems encountered with the introduction of the portfolio of writing into the assessment package for languages was the perception by teachers that it involved much more work than preparing learners for external examinations. It was a new concept for them that learners were being asked to be accountable for their own learning and for their input into their portfolios. Teachers spoonfeed learners a great deal and were taking over much of the responsibility of having neat and complete portfolios of evidence for each and every learner. Time has been needed to help teachers pass the responsibility on to the learners and to be satisfied that an untidy piece of evidence tells its own story and that that is the story that must be told, not a 'doctored' story.

John Abbott (of) says that 'Just because a thing is difficult does not mean that it is not worth doing' and I think that this is the case with portfolios of evidence of coursework. Teachers have had to adjust their assessment styles to meet the demands of formative assessment, giving learners clearer task rubrics and assessment criteria as well as constructive feedback about the different criteria, not merely putting a summative mark at the bottom of a submitted piece of work. Gradually, teachers have relinquished the control of the portfolios somewhat and, after about five years of portfolio inclusion in the languages, it reached a balance of control and input from teacher versus accountability of learner. One of the mechanisms that has been most useful in differentiating the roles of teacher and those of the learners in terms of coursework, has been the introduction of a teacher's portfolio which includes details of all the tasks and assessment instruments used during the year, from which learners will have selected items for their portfolios. These teacher portfolios are crucial to the work of the moderation panel for portfolios at the marking session at the end of each year. The information of what the teacher asked for, the evidence from the cluster minutes as to what the cluster felt about the assessment and the learner's own work together form the basis of the moderation.

Early in its existence, the IEB introduced the User Group concept, where teachers from all schools that write the IEB Senior Certificate examinations could get together regionally and nationally to network and share knowledge, methodologies and experiences. Once a year, a National User Group conference is held, where the previous year's examination is discussed and reflected on; examiners report back to the teachers and outline their vision for the next examination; new ideas for teaching, learning and assessment are explored and workshopped; and proposals on the way forward for the subject are made. This system has been the strength of the IEB and the professional development of teachers that results from this system has fed into the success of the coursework components of IEB assessments. Any innovation or change can be undertaken with some confidence because of the user groups.

The user groups have been the method of 'getting it right' with portfolios of writing in First Languages. Implementing portfolios was not without problems, but the small numbers of schools and the type of language teachers and resources in those schools made it not only possible but effective. However, as the IEB numbers have grown, and the portfolio system has become more widely used across all subjects, so have the problems grown. 'Keeping it right' with portfolio assessment has become more complicated:

- teachers have always 'taught to the test' and have become just as adept at 'teaching to the portfolio requirements'. It is difficult to encourage individuality of tasks when inexperienced or insecure teachers find it more comforting to recycle what other teachers have tried,
- the need to correlate internal and external assessments has arisen – contrary to the rationale for the use of coursework, but necessary for quality assurance and credibility of results,
- more and more checkpoints have had to be introduced to monitor, standardise and moderate portfolio work,
- without formal written examinations, there is always a question mark over who's work is being assessed and the biggest concern about the replacement of examinations with coursework is that of authentication.

The IEB has introduced several mechanisms to combat these concerns:

- the user group system continues to be the mechanism for encouraging innovation, teacher confidence and originality, and seems to be working well,
- the introduction of other components in the portfolio has helped satisfy the correlation concerns,
- the introduction of cluster groups as another standardisation and moderation instrument has been effective,
- the introduction of a Common Assessment Task has made the standardisation and moderation of portfolios easier,
- plagiarism and fraud have increased over the years with the result that stricter control has become necessary.

The Standardisation Policy changed the face of the Writing portfolios. They could no longer be exclusively the vehicle for alternative assessments; they needed to reflect continuous assessment in the areas that **would** be assessed in the external examinations. Thus, 4 tests and the 2 preparatory examinations now had to be included in the portfolio as evidence of this continuous assessment. Teachers have not liked this move. Many feel that the Writing component of the whole curriculum has been downplayed in extending the function of the portfolio and that less time and effort is now expended on the teaching of writing as a result. This perception has been addressed by reviewing the way the portfolio marks are captured, allowing for the Writing section to be reported separately from the continuous assessment (CASS) component. An interesting result of the inclusion of the preparatory examinations in the portfolio has been the realisation by most teachers that schools tend to set stricter examinations for these 'prelims', or 'mock matric' examinations as a way of coercing learners into a final spurt of effort for the final examinations. This strategy may work, but when the

results of the preparatory examinations count, the extra harshness is unfair. Some adjustment has been made and the correlation between preparatory and final examinations is now better.

The introduction of cluster groups was not without problems too. Teachers welcomed the opportunity to meet with other teachers to share concerns and experience, but found it difficult to include cluster meetings into already tight time schedules. They also found that being asked to critique a colleague's standard of assessment was not easy. With trial and error methods mostly, the IEB slowly refined the cluster system until a workable method of clustering could be found. Now, although only two cluster meetings a year are compulsory, many meet more often because they have come to value the networking so much. The first compulsory meeting must take place in the first term of the year and the focus of that meeting is on the teacher's portfolio. Clusters are asked to monitor the degree of difficulty and the nature of the tasks envisaged for the year's coursework programme and to suggest alternatives/adjustments where deemed necessary. The second meeting of the year concentrates on the assessment and looks at the way the teachers have dealt with the CAT in particular and the coursework in general. A useful innovation was the adjustment of the role of the cluster and the forms to be filled in. It is no longer the duty of the cluster to make adjustments to the marking of any member of the group, but merely to alert the IEB to problems of any sort so that these can be dealt with before the final submission date. This has helped greatly and fewer learners are being penalised because an inexperienced/over zealous teacher is too strict for example.

The IEB also drew up and distributed a Moderation Manual to all its schools, which explains in detail all the procedures, time frames, forms etc with which a teacher, a school, and a cluster must comply. More recently, the IEB has introduced a web-based Cluster Report Manager to minimise the paper work and planning and reporting time for cluster meetings. This system is now running well and has helped regulate cluster activity effectively.

The Common Assessment Task (CAT) has several purposes:

- there is a need to standardise the type of tasks presented by teachers of the different languages across the levels of language—all First Language learners complete the same task, no matter in which language and it is the same with Second Language learners;
- the task could encourage teachers to set tasks on unfamiliar areas: it was felt that there were some areas of the syllabus that were not being taught—partly because these were not tested in the external examinations and partly because teachers were unfamiliar with the methodologies or the skills required by these areas (research and reporting on findings is an example of the former and critical language awareness an example of the latter);
- the task could assist moderation of the portfolios: in the first level of moderation (within a school) the HOD could use the CAT as a means of checking whether all teachers in the grade were assessing to the same standard; at cluster level, the group could use the CAT to cross check their schools and report to the IEB if serious problems (inexperience or non-compliance) were discovered—before the final moderation so that the learners are not disadvantaged by teacher inadequacies; and the external committee for moderation of portfolios could use the CAT as an indication of the manner in which teachers have assessed the other components and a check of standards;

- the CAT could be used to introduce different assessment tools and instruments: the use of rubrics, check lists and rating scales has been encouraged to familiarise teachers with these and move towards more criterion referencing than a marking memorandum allows;
- at Second Language level, the CAT has been useful as a means of introducing a body of vocabulary and language structures that are known to have been covered, with a very positive spin off in oral moderation and in the choice of texts for the external papers, because examiners and moderators can use the known to balance what might be unknown to the learners.

The CATs are working well, and are now an entrenched mechanism that has proved and hopefully will continue to prove useful and effective for its purposes. They do assist in 'keeping it right'.

The integrity of the Portfolio process is crucial to its survival. We all know how important matric symbols are to candidates, parents, teachers and schools. The old cliché can be extended to 'All's fair in love and war and exams'. So, any loophole will be eagerly explored by the idle, the desperate or the unprofessional. Authentication is the biggest issue that still plagues coursework In order to combat this pandemic, the IEB has taken several steps:

- it has introduced severe penalties, in line with Umalusi directives, to combat the non-acknowledgement of sources in a learner's work: detected plagiarism becomes and examination irregularity with harsh consequences if proven;
- schools are encouraged to have their own policies on plagiarism and to treat it as seriously as the examination bodies do so that most cases are detected and dealt with before any coursework reaches the final stage of moderation;
- clusters also use their combined expertise to authenticate work, using internet programmes to help detect serious cases of 'lifting' or quoting without acknowledgement;
- most effectively, teachers are encouraged to set tasks that demand a learner's engagement with the issue, so that mere downloading of information does not receive marks: metacognitive and personal response type questions included in the task after the investigation activity, go a long way towards rooting out 'borrowing' of text;
- introduction of forms in the portfolio that learners have to sign, declaring all the work in the portfolio to be their own, is a deterrent to all but the hardened 'borrowers';
- learners, teachers and school principals all have to sign these declarations—so that more careful scrutiny is encouraged;
- more of the tasks in the portfolio have to be completed in class under controlled conditions: final drafts of process tasks, and much of the drafting process and planning activities are monitored by the teacher so that the signing of the declaration becomes easier;
- a review of the manner and timing of the sampling process is underway: informing schools too early of which portfolios are required for cluster or external moderation sometimes results in over-prepared portfolios or interference by teachers/learners; a random sample does not give the moderation panel the full picture of the state of writing in the schools;

- suspected plagiarism that is not provable is nevertheless mentioned in the feedback forms completed by the external moderating committee as a plea for greater vigilance in the future.

The feedback forms have been the most effective mechanism for 'keeping it right' in the portfolios. The moderation panel looks for any discrepancy in standards between one school and the bulk of schools. Where there is not a significant discrepancy, but the standard is not quite in line, the exact problem is outlined on the form and suggestions are made to help the teacher get the standard right. The forms are detailed and the panel members are encouraged to give full and constructive feedback in the interest of the next cohort of learners. Individual teachers can appeal against a perceived misperception of the panel, or request a more detailed discussion/interaction as a result. This process has proved invaluable but could be improved further. The introduction of clearly headed pro-formas to allow for easier and more complete feedback could cover the following categories:

- informing a school that there are serious inadequacies and that the portfolio mark has been discounted. Reasons will be provided and action steps advised;
- informing a school that the portfolio mark is too generous (over marked) and that it will be dropped by a specific percentage;
- a notification that the portfolios and the assessment standards were satisfactory. This would be similar to the form used at present;
- informing a school that the marking was too strict (under marked) and that the centre's marks will be raised by a certain percentage;
- a highly positive feedback form for those schools whose work is seen as exemplary.

To sum up, I believe that the Portfolio approach is educationally sound, is challenging to teachers and candidates, is rigorous and balanced. It combines continuous assessment with formative assessment and puts the responsibility on the teacher to assess and rank each candidates' work accurately. It encourages editing in certain pieces and, through the Common Assessment Task, allows for moderation by cluster groups and by IEB appointed panels. The system has been built up over the years and needs to be cherished.

As we move into the FET phase of the new curriculum, we have to reflect on what changes, if any, we wish to bring to our assessment package. The most obvious and important change will be the use of outcomes and assessment standards instead of content as the focus of the assessment. However, it is not in the coursework components that the difference will be most felt, as formative assessment is outcomes driven, particularly in language teaching and learning. The IEB would like to maintain the two parts of the portfolio: one for the continuous assessment of the full gamut of outcomes and assessment standards that are assessed externally, and one for those types of assessment not best suited to pen and paper tests. The Writing component, comprising extended writing of various genres, as well as the way the fourth genre is treated, lend themselves to coursework. We are confident that our experience in portfolio management and moderation will enable us to continue to keep the portfolio feasible, manageable and reliable enough to be taken seriously by learners, teachers, ourselves as the assessment body and Umalusi as the quality assurer. It is for this reason that

we will be motivating to keep coursework alive and well in the assessment of languages at all levels.

References

- Abbott, J *www.info@21learning.org*
- Handy, C. (1994) *The empty raincoat*. London, UK: Hutchinson
- Selkirk, P - a teacher quoted by Dylan Wiliam in his article in Critical Quarterly (2000) see below
- Wiliam, D. (2000) *Critical Quarterly*, 42(1), pp 105-127

Strategies for answering multiple choice questions among South African learners

What can we learn from TIMSS 2003?

Edith R. Dempster, School of Education & Development, University of KwaZulu-Natal

Abstract

This study investigated the strategies used by South African learners in answering text-only multiple choice items in TIMSS 2003. The investigation focused on 20 items in which learners showed a strong preference for an incorrect answer. The trend was particularly noticeable among learners who attended schools in which all the learners and teachers speak an African language as their home language. The study arose from a concern about the readability of TIMSS items as a valid instrument for assessing the scientific knowledge of learners who lack proficiency in English.

Results are presented for two groups of learners: those who attend 'Africans only' schools, and those who attend mixed-race schools where most of the teachers speak English as a home language ('non-African schools'). These schools previously catered for the Coloured, Indian and white communities.

Sentence complexity was significantly higher in questions where >40% of learners chose an incorrect answer than in questions where learners answered by random guessing, and where >40% of learners chose the correct answer. Only 12 out of 73 items were correctly answered by >40% of learners from 'African' schools, whereas 36 items were correctly answered by >40% of learners attending 'non-African' schools.

Five strategies were detected in items in which >40% of learners attending 'African' schools chose an incorrect answer. The most common strategy (9 items) consisted of favouring the answer that contained the greatest number of familiar words. Learners rejected answers containing words that they did not recognize. Other strategies were selecting answers that contained words that were also in the question (2 items), choosing an answer that contained

all the options (1 item), misunderstanding the question (2 items), and selecting answers that indicate misconceptions (4 items). Two items could not be classified.

The strategies indicated that learners did not understand the twenty questions clearly, and resorted to strategies based on 'tricks' that, with luck, might be the correct answer. Learners attending 'non-African' schools also resorted to 'tricks', although learners were more likely to select the correct answer than the incorrect answer. The results of this analysis raise further concerns about the level of content knowledge among South African Grade 8 learners, and therefore the content validity of TIMSS in South Africa. The lack of appropriate content knowledge is compounded by low levels of proficiency in English, coupled with unclear wording of some TIMSS items.

Introduction

TIMSS 2003 has once again confirmed the poor performance of South African Grade 8 learners in Mathematics and Science at Grade 8 level (Reddy 2006). As with the two previous studies, South African learners achieved the lowest score of all participating countries, including other African countries such as Ghana, Botswana, Tunisia and Morocco. The 2003 study was the first to be conducted with a group of learners who had experienced outcomes-based education throughout their schooling, yet their scores were the same as that achieved in 1998, and lower than the scores achieved in 1995.

In 2003, 8,952 learners participated in the TIMSS study, 11.6% of whom wrote the test in Afrikaans, while the remainder wrote the test in English. Most of those who wrote in English were not home-language speakers of English, with 84% of them being African learners attending schools that served the African population group under the apartheid system of government. The performance of this group was significantly lower than that of learners who attended mixed-race schools that were previously designated for the white, Indian and Coloured population groups (Reddy, 2006). In 2002, the enrolment in 'African' schools was about 96% African, while 'non-African' schools had about 40-50% African learners, and about 50% learners of the race group for which the school was originally established (Reddy 2006).

The universal validity of TIMSS questions has been questioned by a number of researchers, since the content tested may not match the content taught in all the participating countries. TIMSS items are validated by a panel comprising representatives from all participating countries. Items are included if they match the taught curriculum in 70% of the participating countries. In 2003, South African official curriculum was C2005, in which content was deliberately under-specified. Teachers were expected to choose content that would enable them to achieve the learning outcomes, within content areas that were specified only in very broad terms. Thus the content tested in TIMSS may have been invalid for South African schools, since it could vary between schools, and since the official intended curriculum was a deliberately vague document (Reddy 2006).

A scan of TIMSS released items raised our concerns about an additional source of invalidity of TIMSS 2003: the readability of certain multiple choice items, particularly for South African learners. English is the official medium of instruction in most South African schools, including those where all learners and teachers speak one or more of the indigenous African languages. This description applies to 84% of the learners who wrote the TIMSS tests in English. For these learners, English is a second language in the urban schools, and a foreign language in rural schools (Setati et al, 2002), and their scaled science score in TIMSS 2003 was 199, compared with a mean of 483 for learners attending the most resource-rich schools in South Africa, which were previously designated for the white community. The international average scaled score was 474 (Reddy 2006).

A previous study investigated the readability of 73 multiple choice items from the Science section of TIMSS 2003 (Dempster & Reddy in press). The selected items consisted only of text, thus eliminating confounding variables such as visual literacy, or ability to interpret

graphs or tables. We hypothesized that the percentage of learners selecting the correct answer on each item would be negatively correlated with readability factors such as sentence complexity, number of unfamiliar words and number of long words.

Readability formulae such as the Flesch-Kincaid Grade Level Formula and the Fog Formula are only reliable if they are applied to a sample of text containing at least 100 words (Allan et al., 2005), a condition that is generally not met by multiple choice items. Evaluating the readability of multiple choice items is therefore problematic.

Reading research has shown that vocabulary load and syntactic complexity are the most robust predictors of readability. Vocabulary load is measured by the number of syllables, or number of letters in a word, or location on a frequency list. Syntactic complexity is measured by the number of words per sentence, and/or the number of embedded clauses or phrases within the sentence, or the lexical density (number of content words per clause). After testing several models and variables over ten years or more, Homan et al (1994) arrived at the prediction equation that became the Homan-Hewitt Readability Formula. This formula was designed for short sections of text, such as is found in multiple choice questions.

Three factors proved to be good predictors of readability and were used in the Homan-Hewitt readability formula. These factors were:

- sentence complexity, measured as the average number of words per Hunt's T-Unit (WNUM). Homan et al. (1994) defined sentence complexity as the average number of words per clause, where a clause is the shortest grammatically correct sentence into which a piece of writing can be divided. Thus a complex sentence contains one T-unit, whereas a compound sentence may contain two or more T-units.
- number of unfamiliar words, defined as the number of words that were not familiar to 80% of 4th-Grade children (WUNF), as listed in The Living Word Vocabulary (Dale & O'Rourke 1981).
- Number of long words, measured by how many words have seven or more letters (WLON).

The Homan-Hewitt Formula is stated below.

Readability Level = 1.76 + (0.15 x WNUM) + (0.69 x WUNF) – (0.51 x WLON)

The outcome of the formula relates to Grade level for American learners. The formula was validated with a sample of 782 learners with known reading ages, selected from the 2nd to 5th-grade classes in five elementary schools. All the children included in the sample had mastered the subject matter of the test at a level of 75%, thus eliminating content validity as a potential confounding factor (Homan et al. 1994). The readability levels proved to be good predictors of performance, with learners able to correctly answer more questions where the readability level was at or below the child's grade level.

Allan et al. (2005) have criticized the usefulness of readability formulae for examination questions, stating that subject-specific terminology is likely to inflate the readability level,

while it is part of the knowledge required for the subject. They also point out that readability formulae such as Flesch-Kincaid, Fog, and Homan-Hewitt are calibrated for American Grade levels. If used in other countries, they need to be calibrated against the reading competence of children in those countries.

Given the concerns expressed about the applicability of readability formulae to examination questions, and the necessity to calibrate the grade levels to the country, we did not calculate a grade level for TIMSS items (Dempster & Reddy in press).

Since *The Living Word Vocabulary* was developed and standardized for American children, it was not an appropriate word list for South Africa. We were unable to locate a similar index of familiarity for South African learners, and substituted a primary school dictionary for South African learners in its place. The dictionary selected contains a list of 3 300 words that learners should know by the end of primary school, or Grade 7 (Blacquière et al. 1995). The assumption is that Grade 8 learners in South Africa should be familiar with most of the words in this junior dictionary. Sentence complexity and number of long words are not affected by the context of the country or the school, and were calculated in the same way as Homan et al. (1994).

We found that learners attending 'non-African' schools were more likely to randomly guess an answer if the sentence complexity was high, while African learners at African schools were more likely to select an incorrect answer if sentence complexity was high (Dempster & Reddy in press). The average number of unfamiliar words, and the average number of long words per item were not correlated with percentage of learners answering correctly.

An average of 28% of African learners attending African schools selected the correct answer in 73 selected items. This was significantly lower than the average of 39% for learners attending non-African schools. Moreover, learners attending African schools were more likely to select randomly among the answers, and more likely to favour an incorrect answer than learners from non-African schools. In fact, learners attending African schools showed a strong preference for an incorrect answer in 20 of the 73 items, or 27.4% of the items. In these items, 40% or more of the learners selected one of the distractors. By contrast, preference for an incorrect answer occurred in nine items among learners attending non-African schools.

This study seeks explanations for the pattern referred to above. What strategies were learners using in selecting a plausible answer? Can the strategy be explained in terms of the readability of the items? Pollitt and Ahmed (2001) have questioned whether TIMSS assesses science or readability. They analysed several TIMSS items, and showed that learners could apply different forms of logic that led them to an incorrect answer, depending on how they interpreted the question. This study attempts to do the same for 20 TIMSS items in which 40% or more of the South African learners favoured an incorrect answer. The phenomenon occurred more frequently among African learners attending African schools, where the level of proficiency in English is lowest among the range of South African schools. However, it is also apparent among learners attending non-African schools, whose proficiency in English is higher.

Methods

The method of sampling schools, learners, design of test items and administration of the test is described elsewhere (Mullis et al., 2001). A total of 8,952 learners from 192 schools participated in TIMSS 2003 in South Africa. While the total number of learners who wrote the test in English is 7,912, the number of learners per question is smaller because in terms of the TIMSS matrix design, not all learners answered every question. The results presented here represent 6,658 learners attending African schools and 1,254 learners attending non-African schools throughout South Africa.

The study estimated readability using three variables identified by Homan et al. (1994) as the most useful predictors of readability of multiple choice questions: sentence complexity, number of unfamiliar words, and number of long words. Words that were not listed in the Shuters Junior Dictionary for Southern Africa were classified as 'unfamiliar words. Where a word used in a TIMSS item differed from the listed word by more than two letters, it was classified as unfamiliar. Thus for example, 'burn' is listed in the dictionary, but not 'burning', therefore 'burning' is classified as an unfamiliar word. In addition, many words take on different meanings when used in science than in other spheres of life. If the scientific meaning of the word was not given in the dictionary, it was classified as an unfamiliar word. This technique probably under-estimates the number of unfamiliar words for African learners in African schools, but it was the closest available technique in the absence of a familiarity index for South African learners.

Twenty items in which 40% or more of the learners attending African schools selected an incorrect answer were studied to understand how learners may have reached their choice. We also relied on anecdotal evidence gained from students and learners on strategies that they use when confronted with multiple choice questions.

Results

Mean (±SD) sentence complexity for the 20 selected items was 12.7±4.6, mean number of unfamiliar words per item was 4.4±4.5, and mean number of long words was 5.2±3.8. Sentence complexity was significantly higher in questions where >40% of learners chose an incorrect answer than in questions where learners answered by random guessing, and where >40% of learners chose the correct answer. Mean numbers of unfamiliar words and long words did not differ significantly among the three types of answers.

The items were divided into six categories, based on the strategy that best explains the pattern of choice. Examples are given below for released items only.

1. Choosing an answer that contains a term that is also in the question. Two items of the 20 fell into this category.

Example 1:

The burning of fossil fuels has increased the carbon dioxide content of the atmosphere. What is a possible effect that the increased amount of carbon dioxide is likely to have on our planet?

A. A warmer climate
B. A cooler climate
C. Lower relative humidity

D. More ozone in the atmosphere

Sentence complexity of the question is 14.67, it contains 8 unfamiliar words and 12 long words. The correct answer is A.

Table 1: Percentage of learners selecting each answer in example 1

	A	B	C	D	n
African schools	19.6	17.3	12.3	50.8	1019
non-African schools	45.8	8	12.3	34	212

The word *atmosphere* is the only word that occurs in the question and in option D. It was a strong distractor for African learners. Learners attending non-African schools favoured the correct answer, but the second most popular choice was D. Anecdotal evidence from learners and university students indicates that matching terms in the question and in the answers is a widespread practice when students do not understand the question, or do not know the answer. It was also evident in questions where two or more possible answers contained terms that were in the stem, as in example 2.

Example 2:

A balloon filled with helium gas is set free and starts to move upward. Which of the following best explains why the helium balloon moves upward?

A. The density of helium is less than the density of air.
B. The air resistance lifts the balloon up.
C. There is no gravity acting on helium balloons.
D. The wind blows the balloon upward.

Sentence complexity is 8.1, there are 10 unfamiliar words, and 11 long words. The correct answer is A.

Table 2: Percentage of learners selecting each answer in example 2

	A	B	C	D	n
African schools	14.2	37.6	17.9	30.3	1021
non-African schools	40.9	26.1	19.7	13.3	203

The word *helium* appears in the question, and also in answers A and C, but it is an unfamiliar word and answers A and C are rejected by African learners. Option A also contains the unfamiliar word *density*, and is not favoured by the African learners. Options B and D contain the words *balloon* and *up* or *upwards,* which are also in the question. These were favoured by 67% of the African learners. Learners in non-African schools favoured the correct answer A, but the second-favourite answer was B, indicating that learners applied the same logic as learners in African schools.

2. Choosing the option that contains words that are familiar. Nine items of the 20 were placed in this category.

Example 3:

The sun is an example of which of the following?

A. Comet
B. Planet
C. Galaxy
D. Star

The sentence complexity is 11, and the item contains three unfamiliar words and two long words. The correct answer is D.

Table 3: Percentage of learners selecting each answer in example 3.

	A	B	C	D	n
African schools	5.3	42	12.7	40	1022
Non-African schools	7.6	25.3	9.1	58.1	198

Comet and *galaxy* are unfamiliar words, and few learners selected those answers. Answers B and D were equally attractive to learners in African schools, while over half of the learners in non-African schools selected the correct answer, D. Nevertheless, distractor B was the second most popular answer for this group.

Example 4

Which group of energy sources are ALL renewable?

A. Coal, oil and natural gas
B. Solar, oil and geothermal
C. Wind, solar, and tidal
D. Natural gas, solar and tidal.

Sentence complexity is 8, and there are eight unfamiliar words and 5 long words. The correct answer is C.

Table 4: Percentage of learners selecting each answer in example 4

	A	B	C	D	n
African schools	54.5	10.4	16.5	18.6	510
non-African schools	37.6	8.3	35.8	18.3	109

Renewable, solar, geothermal, and *tidal* are unfamiliar words in this item. Understanding the term *renewable* is key to understanding the question. Clearly, in this item, African learners favoured answer A, which consisted only of familiar words. Among learners attending non-African schools, A was a more popular choice than the correct answer, C.

3. Answers that indicate that learners misunderstood the question. Two items were place in this category.

Example 5:

A small, fast-moving river is in a V-shaped valley on the slope of a mountain. If you follow the river to where it passes through a plain, what will the river most likely look like compared with how it looked on the mountain?

A. Much the same.
B. Deeper and faster.
C. Slower and wider.
D. Straighter.

Sentence complexity is 14.7, and the item contains 4 unfamiliar words and 5 long words. The correct answer is C.

Table 5: Percentage of learners selecting each answer in example 5

	A	B	C	D	n
African schools	12.7	41.8	24.4	21.1	994
non-African schools	8.1	34.4	41.6	15.8	209

It appears that learners attending African schools understood the question as "What will the river look like on the mountain?" The item is very badly phrased, with many qualifiers obscuring the question. Learners attending non-African schools fared somewhat better, with 41.6% selecting the correct answer, but B still attracted 34.4% of the learners.

4. Answers that indicate misconceptions or misunderstanding of the concept. Four items fitted into this category.

Example 6:

Which of the following is NOT a mixture?

A. Smoke
B. Sugar
C. Milk
D. Paint

Sentence complexity is 8, there is one unfamiliar word, and two long words. The correct answer is B.

Table 6: Percentage of learners selecting each answer in example 6

	A	B	C	D	N
African schools	46.4	15.2	26.9	11.5	988
Non-African schools	41.4	22.7	27.1	8.9	203

Clearly option A was the most popular choice for both groups of learners. D was the least popular choice for both groups. It indicates a misunderstanding of the concept of mixtures, which can exist in gaseous form as well as in liquid form.

5. Selecting the answer that contains all the options. One item illustrated this strategy.

Example 7

The fossils that are found in the oldest layers of sedimentary rock were formed from which types of organisms?

A. Only organisms that lived in the sea.

B. Only organisms that lived on land.
C. Only organisms that lived in the air.
D. Organisms that lived on the land, in the sea and in the air.

Sentence complexity is 19 for the question, 7 for option A, 6 for B, 7 for C and 13 for D. The item contains 5 unfamiliar words, and 7 long words. The correct answer is A.

Table 7: Percentage of learners selecting each answer in example 7.

	A	B	C	D	n
African schools	11.3	30.2	12.4	46.1	1030
non-African schools	12.3	25.1	3	59.6	203

The key clue to answering this question is the word *oldest*, and the question assumes that learners know something about the sequential evolution of vertebrates in the history of life. In the absence of this knowledge, selecting D covers all the options, and was more popular among learners attending non-African schools than those attending African schools. The second-favourite selection was B, and can be explained in terms of learners' experience of rocks on land, rather than sedimentary rocks forming in the sea.

6. Unknown explanation.

Two items could not be categorized in any of the above categories. One is an unreleased item, and the other is shown here.

Example 8

Eating leafy vegetables is important for human health. This is because leafy vegetables are a good source of which of the following?

A. Protein
B. Carbohydrates
C. Minerals
D. Fat

Sentence complexity is 11, the item contains nine unfamiliar words and eight long words. The correct answer is C.

Table 8: Percentage of learners selecting each answer in example 8

	A	**B**	**C**	**D**	**N**
African schools	43	28.4	18.4	10.3	1057
non-African schools	47.3	22.4	27.3	2.9	205

Minerals is defined in the Shuters Junior Dictionary in terms of substances extracted from the earth during mining operations. Thus, minerals as part of a healthy diet may be an unfamiliar use of the term for South African learners. However, it is unclear why protein should be such a popular choice for both groups of learners.

Discussion

Content validity is questionable for TIMSS in the South African context, since the science covered in South African schools differs substantially from other countries that participated in TIMSS (Reddy 2006). This clearly affected the way South African learners experienced the TIMSS tests, with many items testing knowledge that South African learners had not acquired through the Natural Sciences curriculum. Lack of content knowledge meant that few questions were analysed and answered from learners' knowledge base, but from an alternative set of strategies. Compounding the lack of subject knowledge is the fact that most of the South African learners were participating in TIMSS with insufficient proficiency in English to enable them to understand all the questions.

The items analysed in this study share a common factor: all of them resulted in learners favouring an incorrect answer because of the strategies they used to select an answer to a multiple choice question. Where learners could not apply their strategies, they resorted to random guessing, which was more prevalent among the learners attending African schools than those attending non-African schools. Only 12 of the 73 items included in this study were correctly answered by more than 40% of the learners attending African schools, while 36 of the 73 items were correctly answered by more than 40% of the learners attending non-African schools.

Pollitt and Ahmed (2001) have pointed out that questions have construct validity only if the learners' minds do what we intend them to do when answering the question. They present evidence that by focusing only on the content words in an item, learners could (and do) select an incorrect answer to an item. Linguists agree that the lexically heavy content words (nouns, verbs, adverbs and adjectives) receive more attention from readers than the smaller grammatical words (Pollitt & Ahmed 2001). Through analyzing the content words and the associations these words may evoke in learners' minds, Pollitt & Ahmed (2001) made predictions about which choices would be favoured by learners. They illustrate their predictions with a number of TIMSS items where the pattern of choice across the four or five alternatives indicates rejection of certain answers, and favouring of the correct answer plus one incorrect answer. The incorrect choices can be explained in terms of Pollitt and Ahmed's predictions.

In the case of South African learners, whose scientific knowledge has a weak match with the knowledge tested in TIMSS, and who have the added disadvantage of writing the test in their second language, the strategies they use often have little to do with making sense of the question. Many African learners attending African schools are multiply disadvantaged by poor resources in their schools and at home, teachers who are not confident or well-qualified in science, and they lack proficiency in English (Reddy 2006). South African learners attending non-African schools generally have better-resourced schools and homes, better-qualified teachers, and better proficiency in English. Despite the unequal provision of resources, the best-performing South African learners (those who attend the ex-white schools) score only slightly above the international average.

Hewitt and Homan (2004) have highlighted the importance of readability in large-scale tests, referring to readability as the forgotten validity variable in standardized test items.. Our

analyses have demonstrated a link between sentence complexity and the ability of learners to select the correct answer. The complexity of the grammatical structure of some TIMSS items is not fully captured in a single variable, which counts only the number of words in each Hunt's T-unit. It does not capture the obscuring of the question within complex sentences, the use of the passive voice, questions phrased in the negative, large numbers of qualifying phrases and clauses, and reliance on prepositions and logical connectives.

Overall, the number of unfamiliar words was not significantly associated with the number of learners answering correctly, but analysis of individual items and learners' preferences demonstrates that unfamiliar words affect learners' choice, because they tend to favour answers that contain familiar words. Where the item gave no linguistic cues to work with, the learners' responses indicated random or near-random choice. This was the case for African schools with 17 of the 73 items, and for non-African schools, 7 items indicated random choice. Occasionally, the strategies result in a correct answer being selected, which then obscures the undeniable fact that most South African learners know very little of the science that is tested in TIMSS. If learners answered questions where they do not know the science by guessing randomly, the chance of choosing the correct answer is 25% (or 20% for questions with five alternatives). However, if learners apply strategies that lead them to the incorrect answer, the percentage selecting the correct answer drops to below 25%. The effect of so many questions where the percentage choosing the correct answer is well below 25% is to reduce the average percentage correct to a level just above guessing—the average percentage correct per item was just 28% for African learners attending African schools. The overall score on TIMSS may have been substantially better if learners had guessed the answers or left out questions they did not understand rather than applying linguistic strategies that led them to an incorrect answer.

This study highlights the importance of careful scrutiny of the wording of items to minimize the kinds of strategies used by South African learners in this study. Familiarity indices for all grades would be extremely helpful in writing textbooks and in setting examination papers that are accessible to all learners. Glossaries and dictionaries containing the terminology required for success at all levels of schooling would help teachers and learners to acquire the vocabulary necessary for acquisition of concepts and skills in terminology-rich subjects such as science. Learners should be directed away from resorting to the kinds of linguistic strategies evidenced in this study, since these strategies obscure the purpose of the test, which is to find out how much science learners know.

References

- Allan, S., McGhee, M. and van Krieken, R. (2005) *Using readability formulae for examination questions*. London: Qualifications and Curriculum Authority.
- Blacquière, A., Hoyle, P. and Thompson, C.I. (1995) *Shuters Junior Dictionary*. Pietermaritzburg: Shuter & Shooter.
- Dale, E. & O'Rourke, J. (198*1) The Living Word Book Vocabulary*. Chicago: World Book International.
- Dempster, E.R. and Reddy, V. In press. 'Item readability and science achievement in the TIMSS (2003) study in South Africa'. In *Journal of Science Education*.

- Hewitt, M.A. and Homan, S.P. (2004) 'Readability level of standardized test items and student performance: the forgotten validity variable'. In *Reading Research and Instruction* 43: pp. 1-16.
- Homan, S.P., Hewitt, M. and Linder, J. (1994) The development and validation of a formula for measuring single-sentence test item readability. In *Journal for Educational Measurement* 31 pp. 349-358.
- Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J. & O'Connor, K.M. (2001) TIMSS Assessment Frameworks and Specifications 2003. Boston College: International Study Center.
- Pollitt, A. and Ahmed, A. (2001) 'Science or Reading? How students think when answering TIMSS questions'. International Association for Educational Assessment Conference paper.
- Reddy, V. (2006) Mathematics and Science Achievement at South African schools in TIMSS 2003. Cape Town: HSRC Press.
- Setati, M., Adler, J., Reed, Y. & Bapoo, A. (2002) Code-switching and other language practices in mathematics, science and English language classrooms in South Africa. In J. Adler & Y. Reed (Eds.), *Challenges of Teacher Development: An Investigation of Take-up in South Africa* (pp. 72-93). Pretoria: Van Schaik.

How are teachers' understandings and practices positioned in discourses of assessment?

Anitha Ramsuran, University of KwaZulu Natal

Abstract

Recent curriculum developments in a number of countries have led to the inclusion in formal assessment regimes (leading to 'high stakes' consequences for students and teachers) of a wider variety of methods of assessment, including 'performance' and 'authentic' assessment (e.g. Romberg 1995, DOE 2002, 2005, Broadfoot 1996). There has been a move away from traditional tests and examinations towards tasks that allow more varied, complex responses. This move has simultaneously increased the complexity of the task of assessing student responses—a task generally undertaken by teachers, who must interpret the texts (written, spoken or behavioural) produced by their students in order to evaluate them. This paper reports on a research in progress undertaken with a cohort of B.Ed Hons students (52 students) enrolled at the University of KwaZulu-Natal for a module titled: *Assessment in Education*. The module is highly theoretical and includes an analysis of assessment policy documents. An initial phase of the study (April 2006) examines teachers' understandings of key discourses in the policy documents through a questionnaire. The second phase of the study (June 2006) examines a selection of teachers' (8 teachers) assessment practices through an analysis of assessment tasks, assessment of student responses and interviews. This study is not an impact study but intends to answer the following questions:

- What is the dominant assessment discourses (for example a standards-based discourse) in South African policy documents?
- How are teachers' understandings and practices positioned in these discourses of assessment?
- What kinds of new discourses do teachers' engage with in the module: Assessment in Education?
- Do teachers' practices and positioning shift after engagement with the module/ policies?

For the purpose of this presentation this paper will respond to the first two questions as the data in the second phase of the study has not yet been analysed. In this paper I want to argue that progressivism has left teachers with a disabling legacy when it comes to assessment. Teachers' need access to an assessment meta-language which is sensitive to variations in learning contexts and individual differences. Unless teachers' understand and internalize the language and discourses of assessment, the progressive ideals of the new assessment reform will not achieve its intended goals.

Introduction

Recent curriculum developments in a number of countries have led to the inclusion in formal assessment regimes (leading to 'high stakes' consequences for students and teachers) of a wider variety of methods of assessment, including 'performance' and 'authentic' assessment (e.g. Romberg 1995, DOE 1995, Broadfoot 1996). There has been a move away from traditional tests and examinations towards tasks that allow more varied, complex responses. This move has simultaneously increased the complexity of the task of assessing student responses—a task generally undertaken by teachers, who must interpret the texts (written, spoken or behavioural) produced by their students in order to evaluate them. South Africa has gone through two periods of major curriculum reform producing curriculum documents called: Curriculum 2005 (C2005) in 1997 and The National Curriculum Statements (NCS) in 2002 for the GET phase and in 2003 for the FET phase.

According to Muller (2004), "assessment during the apartheid era was norm-referenced, summative and aggregative" and was characterised by paper–and pencil tests that emphasised academic exercises and the recall of textbook knowledge (Vandeyar and Killen 2003). In Bernsteinian terms assessment was characterised by a strong collection code, which is exemplified by strong boundaries among subjects and between teacher and learner. Assessment of subjects was done independently of each other. Assessment was generally separated from instruction and largely took the form of assessing isolated or fragmented knowledge and skills.

Progressivist pedagogies underpinning C2005 have sought to operate with a 'contentless curriculum'—celebrating individual differences whilst excluding most students from the powerful discourses necessary for success in the wider socio-cultural environment. In Curriculum 2005, the progressive classroom is conceived as operating with a weak classification of subject disciplines through its integrated, 'child centred' curriculum, and an apparently weak classification of the boundaries between teacher (or transmitter of knowledge) and pupil (as acquirer of knowledge)(Bourne 2003: 498).

The NCS released in May 2002 introduced several important changes to proposed assessment practices in schools. For each grade of schooling, there is a set of assessment standards that define the levels of knowledge, skills and attitudes that learners will be required to demonstrate as evidence that they have achieved each phase outcome to an appropriate depth and breadth (conceptual coherence). This means that in each Phase the outcomes remain the same from grade to grade while assessment standards change from grade to grade. The NCS—a streamlined and strengthened version of C2005 has retained its

progressive ideals for example a shift from assessing product to process; norm to criterion referenced; summative to diagnostic assessment.

How have teachers responded to the assessment reform? In an international comparative perspective on outcomes-based assessment, Combrink (2003) qualitatively assessed the realities and problems related to outcomes-based assessment from an international perspective by interviewing a sample of officials in Australia, New Zealand and America and his findings indicate that a major problem in all three countries was a lack of in-service training or inadequate training which made teachers feel incompetent. Teachers have been reported to lack guidance and /or skills –real or perceived-to "implement classroom assessment (Pennycuick 1990, Brown 1991, Bazzini 1993, Emery 1997 as cited in Lubisi, 2000). Another theme that has emerged from literature is that the new assessment increases the workload of teacher (Combrink 2003, Jansen 1999) and that teachers practices may be influenced by time –real or perceived- it takes to design, grade and /or record results of classroom assessment (Pennycuick 1990, Broadfoot et al 1991, Emery 1997 as cited in Lubisi 2000). According to Torrance (1995), considerable resentment was expressed by teachers in the UK at the extra workload involved in assessment. This study intends to understand how teachers position themselves in the discourses of assessment and argue that teachers need access to the language and discourses in assessment for the progressive ideals of the new curriculum to be realized.

Theoretical framework

The analysis of data is framed by a post-structural notion of discourse. Discourse is defined as a 'relatively bounded area (sic) of social knowledge' that both constrains and enables how we think about a particular social object or practice (McHoul and Grace, 1993:31). Thus, discourses allow individuals to interpret particular social situations, or phenomena (Pacini-Ketchabaw and Schecter, 2002). Discourses are about what can be said, and thought, but also about who can speak, when, where and with what authority (Ball, 1990). Discourses embody the meaning and use of propositions and words. Thus certain possibilities for thought are constructed. Words are ordered and combined in particular ways and other combinations are displaced or excluded. Discourses get things done, accomplish real tasks, gather authority' (Said 1986:152). Discourse may have the effect of re-distributing 'voice', so that it does not matter what people say or think, only certain voices can be heard as meaningful or authoritative (Ball, 1994:23). Applications of discourse theory allow valuable fine-grained analyses of documents and texts to be undertaken within a broader structural analysis. This approach is also useful in highlighting values and teasing out competing discourses.

Discourse is the key concept of the relationship between power and knowledge (Foucault, 1977). Discourses are, therefore, about what can be said and thought, but also about who can speak, when, where and with what authority. The effects of power narrow the possibilities of discourse and/or obscure its boundaries. Any particular discourse will generate resistance as it encounters competing discourses. Dominant discursive constructions should be given due weight for the ways in which they circulate and the power that they enjoy (Parker, 1997). They exist within what Foucault referred to as regimes of truth, which make any challenges to the 'realities' to which they refer very difficult indeed (Foucault, 1972). A discourse may gain a dominant position over others in the way in which

it is strengthened and launched by individuals and institutions. Teachers in the educational process are in a powerful position to do just this. Schools are sites where discourses are reinforced and challenged (Corson, 1998). This paper reports on how teachers' understandings and practices of assessment are positioned in the official policy discourses and what new discourses are created in teachers' understandings and assessment practices.

Methodology

This paper reports on a research in progress undertaken with a cohort of B.Ed Hons students (52 students) enrolled at the University of KwaZulu Natal for a module titled: *Assessment in Education*. The module is highly theoretical and includes an analysis of assessment policy documents. An initial phase of the study (April 2006) examines teachers' understandings of key discourses in the policy documents through a questionnaire. The second phase of the study (June 2006) examines a selection of teachers' (8 teachers) assessment practices through an analysis of assessment tasks, assessment of student responses and interviews. This study reports mainly on two data collection techniques: policy document analysis and questionnaire data. The policy documents for the NCS (GET and FET phases) were analysed for dominant discourses and the questionnaires were analysed for teachers' understandings are positioned in these discourses. 49 out of the 52 questionnaires distributed were returned showing a response rate of 94%. The teachers are from five learning sites in KwaZulu Natal (Empangeni -6 Durban-8; Kokstad-4; Pietermaritzburg-23; and Dundee-8). 33 teachers are Level 1; 8 level 2; 2 deputy principals and 6 principals. 16 of the participants were male and 33 female. Most of the participants had under 15 years of experience of teaching. The majority of teachers teach in the foundation or intermediate phase. 28 out of the 49 teachers have indicated that their school has an assessment policy that has been adapted from the National Policy guidelines. In the main teachers were minimally or not involved at all in developing the school assessment policy.

Findings and discussion

Among teachers, there is a growing recognition that the lack of explicitness in curriculum policy documents is causing problems in producing satisfactory assessment procedures. The problem is exacerbated by the one week workshops held by the Department of Education to orientate teachers to the new NCS. 16 of the 49 teachers attended these workshops. Teachers' comments on the positive outcomes of the workshop centred around technical aspects of developing learning programmes and lesson plans, recording, emphasis on changing numerical scores by replacing it with rubrics and checklists, methods of assessment, rating learners achievements. Teachers' voiced concerns about information overload:

> too much in a short pace of time; needed more time and follow up workshops; facilitators not well prepared; some people left the workshop more confused; facilitators not clear about some aspects hence they work-shopped for one week only took 3 days- questions were left unanswered; workshop was too fast paced and superficial, no practical opportunity, increased teacher apathy, does not take context into account; assessment tools and rubrics were not clarified etc. The comments help explain why teachers' understanding of the discourses in assessment are positioned in particular ways. Teachers cannot expect to apply assessment principles they do not understand (Vandeyar and Killan, 2003).

Standards based discourse

The NCS introduced several important changes to the proposed assessment practices in schools. The most significant change was a shift from the criterion-referenced assessment that was recommended in C2005 to a form of standards-referenced assessment. In each Learning Area in each Phase of schooling there is a set of outcomes that define what learners are expected to achieve for each grade of schooling, there is a set of *assessment standards* that define the levels of knowledge, skills and attitudes that learners will be required to demonstrate as evidence that they have achieved each phase outcome to an appropriate depth and breadth. This means that in each Phase the outcomes remain the same from grade to grade while assessment standards change from grade to grade. Because these standards are grade specific, they describe how conceptual understanding is meant to progress in each Learning Area (Vandeyar and Killan, 2003).

For the first time in South Africa, teachers are being encouraged to think of outcome attainment as a continuum of possibilities, rather than as a dichotomy. They are no longer being asked to put learners into categories of achieved/not achieved for each outcome. Instead they have to think about how well each learner has achieved each outcome (Vandeyar and Killan, 2003). This is the approach to assessment that had been advocated *by* Killen (2000). The National Codes for reporting learner achievement introduced in the NCS for the GET phase still require learners' performance of each outcome for each Grade to be recorded on a four-point scale *(exceeded, satisfied, partially satisfied* or *not satisfied)*. This band has been extended for the FET phase as shown in the table below:

Scale of achievement for the National Curriculum Statement Grades 10-12 (General)

Rating Code	**Description of Competence**	**Marks** (%)
6	Outstanding	80-100
5	Meritorious	60- 79
4	Satisfactory	50-59
3	Adequate	40-49
2	Partial	30-39
I	Inadequate	0-29

This reluctance to move to a fully standards-referenced approach to assessment is also reflected in the requirement that reporting should also be normative and "... contain comments on the learners' performance in relation to peers" (Department of Education, 2000, 100). In addition, the tension between progressivism and accountability has negatively facilitated the imposition of external standardized tests for e.g. the grade 9 CTA's which are incompatible both with the classroom experiences of the vast majority of teachers and their learners and the goals of the curriculum. From the questionnaire data, 17 of the 49 teachers did not respond to the question on what they understood by standards based assessment, 3

of these were principals of schools. Multiple understandings, some of which were flawed, emerged from the questionnaire data for example,

- to assess a learner that can be equal to other learners
- level at which assessment should be practised
- process of identifying, gathering and interpreting info about the learners achievements
- based on certain judgement like when a learner is able to do something
- setting an assessment according to level of learners- from their own background or an assessment based on what they are familiar with
- assessment designed by the dept of ed for more than one school e.g., common tests for grade 12 (response from a principal)
- assess learners on what you intended to assess on supervised and controlled tests
- task is assessed according to certain pre-determined criteria.

Embedded in teachers' understandings of standards are fairness, context validity; and maintaining standards and accountability of the system through common tests and standardised testing. In all the definition above, none of the teachers alluded to *assessment standards* that define the levels of knowledge, skills and attitudes that learners will be required to demonstrate as evidence that they have achieved each phase outcome to an appropriate depth and breadth. The definitions '*task is assessed according to certain pre-determined criteria and Level of performance based on criteria used learners are evaluated according to a set standard of criteria'* clearly indicates that some teachers are positioned in a criterion referenced discourse and have not internalised the policy shift made from criterion referenced to standard referenced assessment. For those teachers that were able to offer a position aligned with policy, the definitions fell short in terms of the progressive ideal of setting assessment standards (global competitiveness, standards reference, accountability, transparency) for example:

> assessment based on standards
>
> assessment that is standardised to achieve outcomes that are set, i.e. dynamic assessment
>
> every learning outcome has assessment standards
>
> where you give learners tasks according to their standards
>
> assessment that is based on the expected level of each child in a particular grade
>
> achievement of learners through the assessment standards
>
> what is required of the learner at the end of the LO
>
> assessment that has standards that move from easy to difficult.

When teachers' understandings are positioned differently from policy then it is unlikely that teachers will be able to apply these assessment principles embedded in the standards based discourse to their practices. Another stark disparity is that at the level of policy documents subject assessment guidelines (DOE, 2005 for example in the Life Sciences) do not mention

the word 'assessment standards' throughout the document but is implicit in the description of competence and skills. The official policy discourse of assessment standards are not reinforced in guidelines that develop from the curriculum statements.

Competency based discourse

C2005 and the RNCS showed a shift from the performance model of education (outgoing curriculum under apartheid) to a competence based model. According to Bernstein (1996) performance model serves primarily economic goals and are considered instrumental. Assessment in this model aims to identify what learners have not acquired, i.e., 'what is missing in the product' (p60) and teacher professionalism is grounded in grading procedures. By contrast, in competence models time is not explicitly punctuated (sequencing and pacing is more flexible), assessment emphasizes what is *present* in the acquirer's products.

Competence modes moves away from the concept of deficit (assessment in terms of what is absent rather than what is present). It emphasizes empowerment, for example in the RNCS statements there exist a dominant discourse of liberal/progressive empowerment in advocating learner-centredness; cultural empowerment through redress and political empowerment through the development of a democratic citizenry. In the assessment section of the RNCS policy there exist explicit competence descriptions that embed the notion of empowerment, for example, *'develop justifiable and responsible positions on the influences of different beliefs, attitudes and values in various communities'* (DOE, 2005- Life Sciences subject statements, p51).

Data from the questionnaire reveals that teachers' understandings of competence are embedded in the performance model- with a strong norm-referenced discourse emerging, for example,

> to describe some-one comparing with others
>
> to see whether learners meet the required standard in order to progress to the next grade when a teacher is able to describe learners competence against each other in a class e.g., when answering questions asked by the teacher
>
> levels of performance compared to other people e.g., certain learners are better at practical / performance tasks compared to others
>
> the capacity of continuing performance within specified ranges and contexts resulting from integration of various specified outcomes

What are the theoretical tensions that exist in locating assessment in the RNCS within the competence model? Where do predetermined outcomes fit within a competence model of education? The NCS explicitly states

> To assist with benchmarking the Learning Outcomes in Grades 10-12, subject competences have been described to distinguish the grade expectations of what learners must know and be able to achieve (p47- Life Sciences)

Since the outcomes based approach involves setting benchmarks, it inevitably incorporates the concept of deficit, which is contradictory to competence models (Harley and Parker, 1999). With the specification of explicit assessment standards and competence levels, the benchmark has been set and learners are assessed on what they are able to acquire- a deficit notion. This is reflected in teachers' responses:

> someone who meets the level of expectation, someone who has achieved the required level

> to be able to complete or correctly do a task there are competencies of various kinds but an educator can assess two competencies for e,g, taking care of a microscope and drawing an organelle

> describing performance of learners, how it is going, does it need help from teachers or not to do a task

Evaluation discourse

The Assessment Policy Act (1998) defines evaluation as

> The process whereby the information obtained through assessment is interpreted to make judgements about a learner's level of competence. It includes a consideration of learner's attitudes and values. (Assessment Policy Act, 1998).

The official policy discourse of evaluation includes collection of adequate evidence, interpretation, judgement compared a benchmark: assessment standards and competence description and a summative assessment of a learner's worth. It considers assessment of attitudes and behaviors, not only cognitive dimensions of learning. What does this discourse do? It suggests to teachers (or attempts to impose on them) practices and criteria of assessment to be measured against competence descriptions and assessment standards, and provides them with arguments to make sense, justify and explain their practices to pupils, parents, official moderators, and interested others. Following Bernstein (1999), we can say that this discourse attempts to regulate teachers' practices of assessment.

The discourse of evaluation is not unitary but consists of an official discourse and other unofficial discourses (Lerman, 2001). What are other discourses that teachers' draw on in their understanding of evaluation? A strong regulatory discourse linked to monitoring and accountability is evident:

> when a teacher, learner or any stakeholder reflect on what is being done then see where they went wrong, what are their strengths and what is the way forward to achieve their goals

> is the monitoring standards and effectiveness and determining the strengths and weaknesses of learning

Teachers' also draw on deficit discourses where assessment serves purposes of what is missing and needs to be addressed:

> to weigh for assessment of learners

to look at the learners work as a class and give some clues as to how to make improvements

a form of measuring the students performance and the extent to which the task is performed and achieved

reflection on ones performance with the purpose of identifying strengths and weaknesses determining the success or failure of performance and then improving on it

to check the value and shortcomings of assessment by looking at its strong and weak points

testing the learner's work whether it is good or bad in other words watching performance

to look at the learner's work on their performances whether they have done what was required of them or not

In this position, the teachers speak with the voice of an unofficial discourse, adopting the values of a traditional pedagogic discourse (performance model).

Conclusion

The data reveals that the standards based discourse, competency based discourse and the discourse of evaluation as progressive ideals carry inherent tensions in the way it is articulated in policy and teachers' understandings and practices thereof. Contradictory demands and alternate deficit discourses emerge when the teacher for example, uses the resources of a competence model within assessment structures that impose strategies that are more consistent with a performance model. Unless teachers engage with the official policy discourse, have access to such a discourse (through workshops and in-service training) and develop a meta-language around the discourses embedded in the assessment policy documents the progressive ideals of the curriculum will not be realized in practice. It require teachers to engage with analysis of the construction of the discourse; attempting to explain what interests such a discourse serves (or what changes in society give rise to such a discourse) and how such a discourse resonates with their present practice. Questions that need to be grounded in such an engagement are: what social factors promote dominant deficient discourses, what class assumptions underpin these discourses, how can these discourses be de-constructed to see whether it serves to empower or disempower teachers and pupils.

References

- Ball, S.J. (Ed) (1990) Foucault and Education. Disciplines and knowledge. London: Routledge.
- Ball, S.J. (1990) Politics and Policy Making in Education: Explorations in Policy Sociology. London: Routledge.
- Bernstein, B. (1996) Pedagogy, Symbolic Control and Identity. London: Taylor and Francis.
- Broadfoot, P.M. (1996) *Education, Assessment and Society. A Sociological Analysis*. Buckingham: Open University Press.

- Bourne, J. (2003) "Vertical Discourse: the role of the teacher in the transmission and acquisition of decontextualised language." In *European Educational Research Journal*, Volume 2, Number 4, 2003.
- Combrink, M. (2003) "An international comparative perspective on outcomes-based assessment: Implications for South Africa." In *Perspectives in Education*, Volume 21(1), March 2003.
- Corson, D. (1998) *Changing Education for Diversity*. Philadelphia: Open University Press.
- Department of Education. (2000) Report of the Review Committee - Executive summary, 31 May. Pretoria: Department of Education.
- Department of Education. (2002) Revised National Curriculum Statement R-9 (Schools). Pretoria: Department of Education.
- Department of Education. (2005) National Curriculum Statement FET (Schools). Pretoria: Department of Education
- Foucault, M. (1972) *The archaeology of knowledge* (A.M. Sheridan Smith, Trans.). New York: Pantheon.
- Foucault, M. (1977) *Discipline and Punish: the birth of the prison*. New York: Vintage Books.
- Harley, K. & Parker, B. (1999) "Integrating Differences: Implications of an Outcomes-based National Qualifications Framework for Roles and Competencies of Teachers." In Jansen, J and Christie P, (Eds), *Changing Curriculum: Studies on Outcomes-based Education in South Africa*. Kenwyn: Juta & Co Ltd.
- Killen, R. (2000) 'Standards-referenced assessment: Linking outcomes, assessment and reporting'. Keynote address at the Annual Conference of the Association for the Study of Evaluation in Education in Southern Africa, Port Elizabeth, South Africa, 26-29 September.
- Lerman, S. (2001) A review of research perspectives on mathematics teacher education. In: F-L Lin & T. Cooney (Eds) *Making Sense of Mathematics Teacher Education*. Dordrecht, Kluwer.
- Lubisi, R.C. (2000) "An Investigation Into Mathematics Teachers' Perceptions and Practices of Classroom Assessment in South African Lower Secondary Schools." PhD Thesis, School of Education: University of Nottingham.
- McHoul, A. & Grace, W. (1993) A Foucault Primer: discourse, power and the subject. London: UCL Press.
- Muller, J. (2004) "Assessment, qualifications and the National Qualifications Framework in South African Schooling." In Chisholm, L (Ed) *Changing Class, Education and Social Change in Post – Apartheid South Africa*. Cape Town: HSRC Press.
- Pacini-Ketchabaw, V. & Schecter, S. (2002) 'Engaging the discourse of diversity: Educator's frameworks for working with linguistic and cultural difference'. In *Contemporary Issues in Early Childhood*, 3(3), 400-414.
- Parker, I. (Ed.) (1997). Social constructionism, discourse and realism. Sage: London.
- Republic of South Africa: Assessment Policy Act (1998). Assessment Policy in the general education and training band, GRADES R TO 9 AND ABET Government Gazette Vol. 402, No. 19640, 23 December 1998
- Romberg, T. A. (Ed). (1995) *Reform in School Mathematics and Authentic Assessment* (New York, SUNY Press).
- Said, E. (1986). Foucault and the imagination of power. In D. Hoy, (ed), *Foucault: A Critical Reader*. Oxford: Blackwell.
- Vandeyar, S. and Killen, R. (2003) "Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised National Curriculum Statement hold?" In Perspectives in Education, Volume 21(1), March 2003.

Lessons learnt

From the evaluation of the examination bodies that assess the Senior Certificate

Chris Nyangintsimbi, Umalusi

Abstract

Umalusi embarked on the evaluation of examination systems of the public and private examining bodies that assess the Senior Certificate in April to May 2005. Senior Certificate is a high stake qualification that marks the end of twelve years of schooling and it opens doors to higher education for those who pass with endorsement and to further learning and employment for those who pass without endorsement. The purpose of the evaluation was to establish a national benchmark for the examination systems; gain an insight into the quality management of examinations and the extent in which the system effectively deals with irregularities as they surface at any stage of the examination process.

One of the key challenges, which Umalusi faces as a national quality assurance body for general and further education is to protect the credibility of the certificates it issues. The general public has recently voiced serious reservations about the decline of academic standards for the Senior Certificate. Umalusi responded through various ways to confirm or deny this perception, for example, it undertook research to evaluate the standard of examination question papers of a sample of subjects over a period of ten years, has improved the moderation criteria for examination question papers.

It became critical to evaluate the examination systems to determine their capacity to consistently deliver credible assessment products. Such an exercise would also serve to establish a benchmark for the assessment systems, which will inform plans for future development through the identification of pockets of excellence and devise mechanisms for replicating these and also expose weaknesses that may compromise the integrity of assessment results.

While the intentions for embarking on this project were noble but there were unforeseen revelations, which emanated from the implementation of the project. It is necessary to share such experiences with experts in the field of assessment with the intention of soliciting constructive comments that can inform the improvement of Umalusi's quality assurance processes and hopefully make an impact on the improvement of quality education.

Introduction

The Senior Certificate is a high stake examination that marks the end of twelve years of schooling. It affords those who receive the Senior Certificate with endorsement an opportunity to access degree programmes in higher education while those whose certificates are without endorsement may access diploma programmes in higher education if they meet specific requirements. The majority of learners may either proceed to Further Education and Training colleges or opt for employment. The Senior Certificate Examinations are used for accountability purposes for learners, institutions, administrators and even political officials. The entire country comes to a standstill when the Senior Certificate results are released with various stakeholders expressing their opinions about the maintenance or decline of academic standards and manipulation of results to give a 'false' impression about the health and performance of the education system as a whole.

Umalusi, Council for Quality Assurance in the General and Further Education and Training bands is legally mandated to provide the quality assurance of the Senior Certificate Examinations and to certify the candidates who meet the requirements of the basic and yet very important national qualification. When the pass rates improve Umalusi receives blame from various stakeholders that centre on the lowering of academic standards. Umalusi, as a statutory council of the Minister of Education has to listen carefully to the concerns of the stakeholders and respond appropriately.

Umalusi responded to the public concerns on the lowering of academic standards by commissioning research, in 2004, which looked at six subjects over a period of ten years (1992 to 2002) focussing on the maintenance/ variation of the standard of examination question papers. The research report revealed that the lowering of standards did take place in some subjects while the standard was raised in other subjects but the quality of marking was not appropriate as most of the markers were found to be unable to pass professional judgment.

It must be noted that a number of quality assurance interventions were happening in the national education and training system, which aimed at improving the quality of education. Examples of the quality assurance projects were the development of a new curriculum statement for both the general and further education and training bands, targeted support for poor performing schools, implementation of institutional evaluations to assist schools in the development of their quality management systems, development of The National Policy on the Conduct, Administration and Management of the Assessment of the Senior Certificate and many more.

Umalusi decided to evaluate the examination systems of the Senior Certificate in both public and private examination bodies. Public examination bodies include nine provincial examination bodies and the Chief Directorate: Educational Measurement, Assessment and Public Examinations plus two private examination bodies. The size of the national examination system is about 770 000 candidates that sat for the Senior Certificate Examinations in November 2004. Evaluation was undertaken in April to May 2005.

The purpose of the evaluation

The main purposes of the evaluation were to:

- Determine the degree of compliance with the national policy on the management of the Senior Certificate Examinations;
- Assess the quality management of examination processes;
- Establish a national benchmark for the management of the examinations;
- Pilot the suitability and adequacy of the draft criteria on evaluation and accreditation of examination bodies.

Section 16(6) of the General and Further Education and Training Quality Assurance Act, number 58 of 2001 provides for the accreditation of examination bodies on criteria approved by the Minister of education.

Approach to evaluation

The evaluation teams were constituted out of members of the staff of Umalusi and peers from public and private examination bodies. The sampling of sites was delegated to examination bodies but it had to be based on the criteria provided. Public examination bodies are structured, mainly, into three levels while private ones have two levels. Umalusi visited all head offices and a sample of districts/ regions and examination centres.

Items on the structured interview schedules were designed to utilise the responses received from the lower levels to verify the information provided by officials at head office level. Further verification of information came from the responses of junior staff members at head office and from examiners, markers of examination scripts and internal moderators of examination question papers, examination scripts and internal assessments.

The evaluation process entailed interviewing officials, full-time and part-time staff, inspecting buildings and facilities, tracking processes and evaluating the documented evidence provided. The main approach involved the tracing of the entire examination process beginning with the setting of papers and ending with resulting. Business Economics standard grade was used as an example and it was chosen because it is offered by all examination bodies and is offered by significant numbers of candidates.

Lessons learnt from the evaluation process

Detailed planning is critical

We learnt that is critical to know upfront what one is going to report on and structure the evaluation instrument in a way that will gather the required information. Planning did not interrogate what was supposed to appear on the report and as a result report writing became a problem and it took longer than it was anticipate to write the report as some of the important information was not gathered and had to be collected through follow up requests.

It is important to estimate accurately the amount of effort that would be required to execute such a huge project successfully. Such an accurate estimation must include the human and financial resources, logistical arrangements, skills and time that would be required. There was

no time set aside for contingency planning and the budget was also not flexible enough to accommodate unforeseen expenses. It always takes longer and costs more than anticipated.

Lay the foundation well

The time frames, which were set aside for the conceptualisation, planning and implementation of the project were very tight and, as a result everything done hurriedly and yet Umalusi was undertaking this project for the first time. Rushed work does not add value and first impressions last. The evaluation instrument was not comprehensive enough to collect all the required data and it did not allow for the manipulation of data. It is critical to have sufficient time for training evaluators and also to deliver an effective training programme. Sampling of sites ended up being delegated to examination bodies to do but based on criteria determined by Umalusi. There was a need for piloting before the main project was rolled out. However, the healthy relationships between Umalusi and the examination bodies made a daunting task to be manageable. The relationships are based on trust, openness and honesty.

The South African examination system needs a common standard/ benchmark

Vast disparities exist among various examination bodies with regard to the levels of quality in terms of:

- Strategic leadership and capacity;
- Systems development;
- Quality management of examination processes, procedures, review and improvement;
- Resource allocations in terms of financial, physical facilities, technology and particularly human resources; and
- Levels of expertise.

Disparities are particularly evident among provincial examination systems especially between those who inherited systems of the previous administrations and newly created provincial departments. The newly created provincial departments are also burdened with the challenges of integrating the former homeland administrations, which were poorly established in the first place and they deal with big student numbers and are operating in geographical environments whose economic growth is declining. However, it is amazing to realise the level of commitment displayed by such examination bodies who, out of meagre resources, still administer successful examinations.

Quality management

Well-planned, documented, monitored and tightly coordinated and reviewed processes result in an efficient and credible examination system. Undertaking risk assessment and management of risks in the system ensures sustainability and consistency. However, it was observed that:

- Very few policies, processes and procedures were adequately documented;
- Monitoring of processes was inadequate and monitoring reports, where they exist, did not always feed into improvement;
- Many have coordinating structures but these were not effectively managed;
- Very few assess and manage risks; and

- Review does not occur regularly and when it does it seldom leads to improvement.

However, the existence of a nucleus of experienced officials lead to successful implementation but this is not sustainable as the majority of these is close to retirement.

Policy development

The level of policy development and compliance determine the examination body's capacity to conduct a credible examination, that is, handle examination irregularities and deal with offenders effectively. In general standards in this regard:

- Vary from excellent to inadequate;
- Few have regulations or documented policies and procedures;
- Few have fully functional examination boards;
- Processing of cases of misconduct that are linked to examination irregularities take very long before they are concluded due to insufficient evidence.

Weak feedback loop

Very limited feedback is given to either the schools, those who perform examination related functions or to examination bodies on the standard and quality of performance, for example:

- The review of policies, processes and procedures varies from adequate to insufficiency;
- Interrogation of examination products in relation to how they effectively assess the curricula is inadequate;
- Implications for teaching based on the findings of each examination cycle does not occur in most cases;
- Implications for curriculum review and development are not analysed;
- Training for examiners and internal moderators is inadequate and consequently, most examination question papers do not meet the required standards and where they do the standard of marking is inappropriate;
- The setting and internal moderation of examination question papers, internal assessments and marking are monitored, chiefly, for compliance with submission dates and not for quality products.

As strong value system carries the day

Examination, like education in general, is a value driven undertaking and this observed through:

- High degree of integrity among the majority of staff and officials throughout the national examination system. This has and still is the backbone of the credibility in the system.
- Willingness to learn and commitment to service delivery are other pillars of the examination system.

Even the least resourced examination system has elements of best practice

Examples of identified pockets of best practice in a sea of limited resources include:

- Facing challenges and managing them creatively;
- Leadership and strategic management;
- Quality management and coordination;

- Effective performance of a nucleus of experienced employees;
- Policy development and implementation;
- Archiving and storage of classified information;
- Systems for improving academic standards of examination products within the constraints of existing capacities;
- Securing the entire examination process;
- Strong belief in values that underpin integrity;
- Mentoring and development of new markers.

Conclusion

The objectives of the evaluation were met and valuable lessons were learnt. Umalusi is aware that the national examination system has mastered the administration of examinations in that very few serious irregularities occur in any examination cycle and the policy stipulates unambiguously as to how these should be handled. The critical challenge lies in delivering good examination products consistently across various subjects and also from year to year. There is a need for developing a comprehensive and effective training programme for those who are charged with the delivery of examination products.

Negative aspects of high stake examinations have their ripple effects in the system in that poor performing schools do not receive adequate support and consequently they do not effectively get out of the cycle of poor performance. Most of these schools are located in the pockets of poverty in the metropolitan cities and in the rural areas. There is a need to provide more resources to newly created examination bodies and to support them in the development of their examination systems

Openness and cooperation made a difficult task manageable and Umalusi is indebted to examination bodies on this.

Elements of best practice were shared, interrogated and documented and this will form the initial benchmark. Criteria for evaluation and accreditation of examination bodies were improved based on the evaluation process and they serve as the benchmark, which the examination bodies must work to achieve.

References

- Department of Education (2004) National Policy on the Conduct, Administration and Management of the Assessment of the Senior Certificate, July 2004
- Republic of South Africa (2001) General and Further Education and Training Quality Assurance Act, number 58 of 2001

Does Quality Assurance Improve the Quality of Education?

Lessons from South Africa

Peliwe Lolwana, Umalusi

Abstract

In this paper quality improvements in education through quality assurance interventions, will be explored through the lens of the South African educational transformation project of the last twelve years. In the country's illustrious constitution, quality issues have been enshrined, and quality assurance is the assumed logical modus operandi for this to happen.

Quality assurance concepts are found in the very first education laws and policies enacted at the beginning of the new democracy in South Africa. Assessment, quality assurance of assessments, accreditation and certification issues have occupied a pivotal position in the entire system. In addition, ten years ago South Africa introduced a National Qualifications Framework, which essentially reinforces quality assurance over and over again.

Can South Africa boast of evidenced achievements in the improvement of quality of its education system? This paper uses South Africa as a case in point to illustrate the complexity of the education industry for quality assurance to make claims in improving the quality of education. At the same time, this paper will attempt to offer some sober suggestions about the appropriate ways of operationalising quality assurance in an education system.

Introduction

For a long time, the South African education system was characterized by strong features of inequalities and large pockets of poor quality. During the apartheid era, the politics of education were closely intertwined with all political efforts to bring the apartheid government down, as protests over poor education became the order of the day by millions of students who chose to sacrifice their education and put liberation of the country first. It is no small wonder then that the new democratic government has had to prioritise the agenda of quality education to all its citizens. Quality education is enshrined as a right[1], in the South African Constitution, that is to be expected by its citizens. The South African Human Rights Commission is currently taking up the issue of the realisation of this right as part of its human rights investigation. It is clear that South Africans put primacy on quality education, but the question to be asked here is what that means and how the country has gone about to engender quality in the system during the last twelve years.

[1] Chapter 2 of the Constitution of the Republic of South Africa Act 108 of 1996

South Africa has the trappings of a first world country, save for the fact that the majority of its people are still locked in extreme poverty and the wealth is in the hands of only a few, hence the South African colloquial is that of a 'first and second economy'. The education system has a schooling component where ±90% of young people attempt to finish twelve years of schooling. There is one externally set examination which certifies successful completion of schooling and also doubles up as a higher education admission certificate. Government statistics show that in 2002 almost 12 million learners were enrolled in the public school system (Department of Education 2003). The private school system is smaller and tends to be elite, with its roots from the earlier missionary schools, but there are some schools of dubious character which have emerged in the last few years especially in high density areas where immigrants and people from rural and outlying areas have come to settle. Both the public and private school systems are controlled by the South African Schools Act[2], which requires all to write the 'same' end of school examinations.

Parallel to the school system is an adult education system, whose enrolments tend to be bigger at the levels parallel to the primary schooling levels. A relatively high number of South Africans still cannot read and write, and a higher number can only be considered functionally literate, having completed between five and seven years of schooling, (Schindler, 2004). There are also many adults who continue to strive to improve their educational attainments, outside of the schooling system. The same Senior Certificate examination written by school-going learners, is a highly sought after qualification by adults as it serves as a gateway to higher education and employment possibilities. However, almost a third of all young people who write the senior certificate examination are young adults who are not in school who may be making a repeat attempt or may be coming back for a second chance education.

Another system which runs parallel to the senior secondary system levels is that of technical and vocational education. There are fifty mega public further education and training colleges, each with multiple campuses, which cater primarily for young adults. Private vocational education exists but the range of provision tends to be narrower and in niche areas. Another large provision is that which exists in industries as the country has a legislated framework that forces companies to invest in training of their personnel, and new labour market entrants (Skills Development Act, 1998). Then there is the higher education system. The focus of this paper will be limited only to the pre-tertiary levels of education, and this will be further curtailed to the general and vocational tracks in the public education system, where the majority of learners are. Although the technical and vocational track can be supplemented by industry training to make the individual occupationally proficient or employment ready, this aspect of training is not meant to be the subject of this paper.

More than a decade of transformation in the education system

The South African education system is a totally different animal from what it was twelve years ago. After inheriting a poor education system for the majority of the learners, the democratically elected government, representing the aspirations of most South Africans, has

[2] South African Schools Act (1996)

worked very hard in transforming its education system. The preceding apartheid government was characterized mainly by separate and unequal treatment of 'citizens[3] on racial basis. The inequality formula permeated all government resources distribution in education; this resulted in *access* and *quality* issues being the most important radars for propelling educational transformation for the new government.

Early policy documents constructed in preparation for the new government are a testimony of this focus on expanding access to educational opportunities; and in fact in these documents the tension between access and development or quality was overplayed (National Policy Education Investigation, 1992). The new government had to deal with a range of factors, contributing to some of the extreme cases of school ineffectiveness, at the beginning. It took over a schooling system that had collapsed in many parts of the country. In 1992 alone, there were more than 16 000 instances of disruptions that took place throughout the country. These disturbances were not caused by students alone, but teachers often staged sit-ins and long term strikes, which contributed significantly to the erosion of the culture of learning in the schooling system (Chisholm & Kgobe, 2003; Chisholm, 2003). The desegregation of the school system, which was characterised by very low educational expenditure, high proportions of unqualified teachers, a high student-teacher ratio in black schools, had manifested itself in extremely poor end of schooling results as compared to those of other racial groups.

Table 1: Grade 12 Examination Results, 1993

	Total Passes In %	Qualifying for University Entrance in %
African	**39**	**8**
Coloured	**86**	**21**
Indian	**93**	**45**
White	**95**	**42**

Source: Mkwanazi-Twala, Mwirira & Greenstein (2003)

To dismantle apartheid education, the new government had to address a lot of issues in the education system in order to afford all learners equality to educational opportunities. The very first focus of the new government has been on educational policies to dismantle segregation in the system, to move towards closing the gap in the allocation of resources in all schools, to return the culture of learning in education, to establish a democratic governance system that allows communities and parents to participate in the affairs of their schools, to move away from what was perceived as 'poor and rote learning curriculum' to a 'richer outcomes-based' education; and for the first time considerable efforts were put into the establishment of a National Qualifications Framework which would focus the energies of the nation to the establishment of standards in all education and integrating education and training in one system.

The transformation of the South African education system has been relatively successful, in many of the affected areas as evidenced by the progress made in the past twelve years in

[3] Africans Were Not Given A Citizenship Status in the Country of Their Birth Though.

education (Chisholm & Kgobe, 1993; Motala & Tikly, 1993; Greenstein & Mkwanazi, 1994; Chisholm, 1994; Tikly & Motala, 1994). In the first few years the new government's reform focus was to integrate formerly divided bureaucracies across diverse racial and economic conditions, and large geographic areas. This was achieved in a short space of time (Department of Education, 2001; 2004). In solving conflict in schools, the new government established a Culture of Learning and Teaching campaign, which had relative successes in normalizing the situation in schools all over the country. The primary school Net Enrolments Rate (NER) has improved from 92% in 1991 to 95% in 2001. The greatest improvement of the NER is noted in the first few years of the new government, and especially during the period 1997 to 2001 where the improvement is recorded at 23% (Education Foundation Trust, 2005). At the secondary level the NER improvements have also been noted. From 1994 the NER for secondary schools improved from 56% to 61%. South African teacher education scenario began to take a new shape, with issues of rationalisation, appraisals, development and support being discussed at various levels. Concrete accomplishments were evidenced by drastic moves to reduce unqualified teachers in the system, whilst the most comprehensive framework for teacher development has been established, which accepts teachers as the essential resource in the education system (Department of Education, 2005). It is only toward the end of the first decade of the new government that quality and standards issues have begun to surface as a major concern in the South African education system.

Quality concerns in the South African public education system.

Quality and standards concepts in education are complex and often span a continuum, which starts from access and infrastructural concerns to relevance and standards issues. In its transformation trajectory, South Africa has experienced the dilemmas of having to address varying quality needs in its attempt to grapple with quality concerns in education, as the question of standards is not a simple one, because standards mean different things to different people. This dilemma is poignantly pointed out in Umalusi report (2004):

> (standards) can refer to benchmarked norms and standards of assessment, statistical standardisation procedures, and curriculum standards defined as part of the curriculum. More popularly, standards refer to a commonly accepted level of performance. However, these standards vary across space and time, including those which are commonly accepted. A high standard for one person may be a low standard for another. What was one time considered a high standard may in another be considered a low standard. Who defines standards for whom is another critical question. Standards are defined in the context of social and cultural norms, which themselves are considered appropriate by some and inappropriate by others. The influence of context in the definition and achievement of standards is also crucial (p.2).

However, in the South African education context, the standards of the general education output continue to be put under the spotlight. The single and only indicator with a 'bite', in the school system, the results of the Senior Certificate examinations (grade 12 results) has served mainly as an accountability tool which holds schools in check but may not be sufficient for quality improvements in the whole system (Taylor, Muller & Vinjevold, 2003). The pass rate for the senior certificate has been steadily rising from a low of 47% in 1997 to a high of 73% in 2003, an improvement of 26% over a six-period (Umalusi, 2004). The

significant increase in pass rates triggered questions about the standards in this important examination and the quality of education in general, especially in the context of a transforming system.

The questioning of standards in the South African education system also has a history that is closely linked to the country's racial divisions. There has always been an implicit assumption that whites received superior education to their black counterparts. Allais (2005), points out that this assumption was vigorously challenged in the 1980s when it was argued that although white standards were seen to be accepted as a benchmark, in many instances, the school ending examination was still geared toward the memorisation of facts. This assumption was also rejected because it was seen in the same light as the assertion that 'all that is white is best and that which is black is worst' and in the political environment of the 1980s this notion was not acceptable. It was not acceptable because the distribution of resources disadvantaged black students and schools, and this had to be put right first in order that all could compete on the same footing. But this debate was also useful as it opened up a space to talk about what was inherently wrong about standards in the schooling system. Notwithstanding the ambiguity about the concept of standards, standards still remain as the mostly referenced point when the public evaluates the quality of the education system. In other words, it is a contradiction in terms to talk about improvement of quality in the absence or perceived low standards.

Another important question that soon arose in relation to the quality of education is that of *relevance*. This, in South Africa, is seen as relevance of education to the needs of the economy and individuals' vocational aspirations, as well as broader social and cultural values (Department of Education, 2000). All end of school qualifications are often evaluated in relation to their ability to meet the requirement of the labour market. However, the South African labour market has been particularly tardy in absorbing school graduates, and many economics and labour market analysts have been at pains in explaining how this rapidly growing economy is unable to produce jobs for school leavers (Bhorat, 2005).

Nevertheless, the education system has not escaped unscathed as the public easily attributes this labour market aberration to the quality and irrelevance of educational outputs. It is for this reason therefore that vocational education has received a lot of attention as it is seen as the pathway with a greater chance of making secondary education relevant and improves the quality. Many critics have already condemned the soon to be introduced vocational education curriculum as being an "abdication of responsibility (by the Department of Education)…. and out kilter with broad state policy on skills acquisition" (Mail and Guardian, 2006). Everybody wants vocational education, think that it will be a panacea for employment and quality, but there is no agreement on what form of vocational education will make this learning pathway to be of good quality whilst making education relevant.

In almost all educational reforms, there is a direct link between either the introduction or review of vocational education for most countries' educational transformation agendas. In spite of this, it is apparent that there is very little common in understanding of vocational education at the secondary level, all over the world, (Young 2005). Because vocational education should be more closely linked to occupations and thus employment opportunities

than, say general academic education, it brings other dimensions as pre-conditions for its success or failure. The different waves of educational reforms also bring along different approaches for reforming vocational education. Educational reforms that affect vocational education often reflect a mixture of planned and unplanned activities, which characterise the state or labour markets' response to perceived needs. These responses range from skills development, strategies for addressing unemployment and the social problems of the young, nationalising qualifications, and the creation of an alternate route to higher education. These responses often culminate in different and often conflicting approaches of constructing and pursuing vocational education, and sadly resulting in vocational education being perceived as a poor cousin of general education, (Young, 2003; Allais, 2003).

Until recently there has not been much interest and concern over the quality of vocational education, save for the historical ties this used to have with the apprenticeship system. The Industrial Training Boards (ITBs) and Artisans who worked as trainers, ensured that technical colleges which provided the theoretical aspect of this education system, continue to be relevant and move with the modernisation of curricula. This binary system, which made up the apprenticeship system, always ran parallel at the level higher than the ITBs and colleges, like the government departments. In other words, colleges continued unchecked by the Department of Education on the question of standards, and the ITBs ran the system with little accountability on standards to the Department of Labour. It is therefore ironic that this system, which on hind sight produced the skills needed to build this country through the apprenticeship system, is now seen to be faltering because it has not provided that important 'integration' ingredient by the departments of education and labour. Clearly, the absence of relevance and standards cannot be apportioned to a 'disintegrated' government system as in the past this has never been a hurdle.

To sum up, in South Africa, concerns about the quality of education have arisen and still continue to be topical, out of concerns of the standards of the examinations in the general education track. In the vocational education stream, concerns have been raised in the main about its ability to make graduates employable. As a result, attempts to improve the quality of vocational education, has resulted in a flurry of activities like developing a National Qualifications Framework, a plethora of new vocational qualifications, and new quality assurance approaches have since emerged.

The rise of quality assurance in the South African education system

The goals for quality and standards in the South African public school system have a long and enduring history. The major instrument used to keep standards in check has always been the Senior Secondary Certificate Examination (SCE), taken at the end of twelve years of schooling and this still remain. Because of the Universities powers to use the results of this examination to select and admit learners to higher education, this examination continues to control the whole school curriculum. Trümpelman (1991) notes that in the history of the SCE, there was never a time when the public was satisfied with the standards of this examination and criticisms were constantly levelled at the various examinations throughout the entire reign of the Joint Matriculation Board (JMB), (1918 – 1986), as well as the subsequent provincial examinations boards. Until recently there has not been much interest

and concern over the quality of vocational education, save for the historical ties this used to have with the apprenticeship system. The Industrial Training Boards and Artisans who worked as trainers, ensured that technical colleges which provided the theoretical aspect of this education system, continue to be relevant and move with the modernisation of curricula. It was a system where everyone knew their place in this hierarchy, but not without displeasure.

It must be noted that whilst the stronghold of universities on general education was not necessarily receiving great reviews by both the public, subject specialists, teachers as well as the university sector at large, many other things were also happening in education that made this form of quality control untenable. Higher education itself was going through some significant changes, which resulted in both a massive expansion in the sector as well as its own differentiation of institutions. A vocational higher education institutional type, the Technikon, emerged around the late 1960's through the Advanced Technical Education act of 1967. There was a significant increase in technical offerings of subjects, like business and service oriented studies, which were not for supporting the apprenticeship system. The apprenticeship system was also experiencing a massive downsizing. The senior secondary school system also expanded significantly and served many other purposes other than selection for university. Finally, the voices of those who felt excluded in the judgment of standards were becoming louder, especially those who claimed to be speaking for industry.

When the legacy of an apartheid education was overhauled, all dissatisfactions about standards and in particular, who controlled them and how they get to be decided, came strongly to the fore. The democratisation of government in South Africa came at the same time when the notion of National Qualifications Frameworks with their attendant concepts of unit standards, accreditation, assessor training and level descriptors had been gaining more momentum, as a replacement system to the examination mode in vocational education at least, in the United Kingdom and other Anglophone countries (Young, 2004). It was no coincidence therefore that when South Africa looked around to replace a discredited apartheid system, the notion of a National Qualifications Framework with its quality assurance and unit standards apparatus was very appealing and became the very first intervention to be legalised by the new government. The first law to be put in place in education by the new government was the South African Qualifications Authority Act of 1995, whose primary purpose was to establish a National Qualification Framework (NQF) (South African Qualifications Authority, 1995).

The establishment of a National Qualifications Framework in South Africa brought significant changes in the South African education system. To this end, a range of agencies in charge of quality control have since been established. There are thirty five Sector Education and Training Authorities (SETAs) established through the Skills Development Act of 1998, sponsored by the Department of labour (Skills Development Act, 1998). The department of education has also established two quality assurance agencies. For the higher education sector there is the Council of Higher Education, and for schools, further education and train colleges which offer technical and vocational programs, and adult education centre, Umalusi, is the council which oversees quality. This plethora of institutions overseeing quality and standards in same institutions, has undoubtedly caused a lot of 'turf wars' as institutions are often pitted against quality assurance agencies.

The NQF system was meant to address a number of issues in a system seen to be in need of a complete overhaul. Its objectives were to:

- provide parity between general and vocational qualifications and curriculum;
- provide portability between one site of provision to the next;
- curbing the marginalisation of vocational and occupational qualifications in most education systems;
- establish progression routes to further learning, and in particular higher education;
- promote democracy and greater equality;
- be responsiveness to the labour demands; and
- provide transparency in standards that would result in improved quality in education.

Over the last ten years, the NQF became the dominant discourse for quality and standards in the education system. However the assumptions, on which the NQF, its structures and processes were established, have not gone uncontested. For example, Mokgelane et al (2003) have this to say:

> Some of the assumptions of advocates of the NQF …….can also be questioned. It is doubtful whether education can and should be the key instrument for transforming labour markets and boosting growth rates, or whether it can function as a panacea for economic ills. One would have to go beyond structures of accreditation to deal with the content of education and training to tackle poor or inefficient educational performance. The curricular aspect of the NQF has been neglected in the quest to design a coherent administrative framework, a necessary but insufficient condition for its viability (p.248).

Others also have added their voices on the list of NQF critics (Allais 2003; Young, 2004'2005). In spite of this, the NQF, which is anchored in the standards generation and accreditation discourses, still remains a dominant and highly influential aspect of the quality and standards trajectory. The NQF was constructed to be a tool for defining 'standards', primarily internationally, outside the formal education system. The NQF borrowed heavily from the quality assurance approaches which were emerging, relying heavily on compliance of institutions to a Quality Management System (QMS). A quality assurance approach puts emphasis on the compliance of the 'providing institution', in its administration and management capacity, to a system known as the 'the Quality Management System (QMS)', and the monitoring of institutional assessments of students. It is a model not very dissimilar to approaches adopted in quality assurance models in manufacturing industries. Young (2004) makes an observation that this is not unique to the South African education system as:

> In searching for some alternatives for ensuring quality, education policy makers, initially in he USA but later in most Anglophone countries, turned to industrial practice for their models. Criterion referenced approaches to assessment, quality assurance systems, and unit standards-based definitions of competence all have their origins in methods developed in mass production and the need to guarantee that products which would later be assembled together to make, for example a car, could be relied on to be identical (p.3)

Like it has been in the USA, and later in most Anglophone countries, South Africa's version of quality assurance of education institutions, is embodied in the NQF, in the form of the accreditation of programs and institutions. This model has not been sitting comfortable with the known quality assessment practices of external examinations and inspection in education.

Has quality assurance improved the quality and standards in the South African education?

Although the concepts of quality and standards in education are universal and enduring concepts in education, they are also difficult to come into full grips with as they are relative and immersed in particular contexts. It is for this reason therefore that in our assessment of the question above, we have to go back to the beginnings and answer it in relation to the South African education context. In other words what is it that we were seeking to improve when quality and standards issues began to emerge as a major concern in the system.

When the new government took over, the uppermost concern was to deracialise all education and provide equal access to education for all as soon as possible. The task of deracialising more than 36 000 public schools, and 152 technical colleges is no small feat, and that the new government has succeeded in doing so with minor disruptive incidences is another evidence of the miracle nation South Africa is. In the first place, the South African Constitution, which holds in balance human rights, social justice and the potential of each child, provided a firm base on which to build an education system that changed the social order created by apartheid education. Progress made in this regard has already been elaborated upon.

However, all is not well in the South African education system. Firstly, research shows that the interdependence of economic and social development and education are always going to undermine the distribution of quality education in South Africa. Whilst we have examples of learners who excel in poor neighbourhoods, this still is an exception than a norm, and poor schools by and large are in poor neighbourhoods, and rural African schools are over represented in this category. Secondly, whilst participation rates are high, the flow through of learners in secondary schooling is still a major problem, so are low pass rates for university entrance; so are literacy and numeracy skills in both primary and secondary phases, so are concerns about the new curriculum which underemphasise the importance of content knowledge, so is the language policy and practice which is a second or third language for the majority of learners, and so are the inadequately addressed needs of special education learners; so is the inadequacy of vocational education curricula in addressing the labour market needs (Nelson Mandela Foundation, 2005; HSRC and EPC, 2005; Mail and guardian, 2006), and the list is long.

The very first attempts to transform general education came with Curriculum 2005, and this was a very large project. Curriculum 2005 was typical of first attempts to introduce an outcomes-based education in line with the NQF project - ambitious. It was characterised by no specified content, no subject boundaries, and basically a dramatic departure from what schools were used to. The revised version of C2005 became known as Curriculum 21st century (or C21) and provided an improvement in many of the areas which were found to be problematic in C2005. Also, South Africa has attempted to engender quality assurance in its

school system through a process of Whole School Evaluation (Department of Education, 2002), which is an attempt to resuscitate an inspection function previously done away with during protests against apartheid education.

The intervention efforts we have had so far, done away the racially differentiated system, but reproduced a class differentiated system, where socio economics has become the major determining factor about the quality of education for the schooling system recipients. In other words, the quality of education for the poor still remains poor. Further, it is increasingly becoming clear that the outcomes based curriculum reform we had initially embraced, is highly unlikely to enhance quality in the system. It is unlikely to do so, because standards cannot be defined by outcomes only, away from knowledge and skills, and away from educational institutions as source of authority on qualifications and provision, (Allais, 2003). It is also unlikely because it brings questions about the large bureaucracy of assessments which overshadow teaching and also the lack of trust by the public of predominantly school based assessments. Lastly the Whole School Education project would be very welcome when and if it finally takes off.

According to Allais (2003), this model for quality assurance had enormous appeal in the South African society, as it promised to 'democratise' the education system and wrest power from the few elite universities. Stakeholders could now control the standards of qualifications. The problem of quality assuring education without a road map is more dramatic in technical and vocational education, where qualitative indicators are difficult to formulate. Vocational education, with its historical perceived low status in British colonies (Allais, 2003), was more susceptible to reforms espoused in the National Qualifications Framework as it was hoped that this would bring about much needed equivalencies between academic and vocational education. According to Allais (2003);

> Part of the seductiveness of this idea was a discursive appeal to integration and equality. Social integration, and the creation of single systems where there had been racial and ethnic divisions, was of overriding concern to many South Africans. Similarly, equality for all was the project of the new democratic state. ……. The strong association of the NQF with the broader liberation project, including the goals of the of social integration and redress, has resulted in many assuming that the theoretical position it assumes is sound; in other words, its moral purpose is confused with its mechanisms (p.10).

This confusion is more prevalent in the new vocational qualifications as the integration of education and training is seen as being sacrosanct by many, in spite of a past artisan system which was never integrated.

Also, as the NQF was bringing about a new way of thinking and talking about standards in the South African education and training system, quality and standards language and apparatus began to focus on the setting of standards, meaning the generation of outcome statements for small units of learning. This new language got confused with the notion of standards and quality in education as it was known as 'standards setting', and (Allais, 2005) maintains, and rightly so, that we 'hit a derailment in the standards construction discourse through this confusion'. It was a derailment because this process should never have been regarded as a process of maintaining standards. This problem was exacerbated by the fact

that the concept of quality assurance was so closely tied to this flawed process (SAQA, 2001). We also have been derailed by confusing the terminology of standards setting with that of standards developing (Allais, 2005). In other words, we assumed that the unit standards which are being set in the NQF regime result in standards of high quality in the system.

In the vocational education stream, the reconfiguration of the qualifications, through a standards generation process, has produced a proliferation of qualifications to an unprecedented scale. These vocational qualifications are extraordinarily diverse and many are occupationally specific, as the following examples indicate:

FETC: Fundraising
FETC: Dry Pet Food Advanced processing technology
FETC: Debt Recovery
FETC: Medical Claims Assessing
FETC: Electrical Network Control
FETC: Statutory Intelligence
FETC: Trade Exhibitions
FETC: Beauty (Nail Technology)
FETC: Cigarette Filter Rod Production Technology

Most of these qualifications are new inventions and have with no proven record of their currency and demand. The traditional vocational qualifications were strongly based on links between vocational teachers in colleges, professional bodies and university or Technikon faculties in applied fields such as engineering and business studies and work-based apprenticeships. In other words the curriculum was based on disciplinary knowledge, same as in schools selected according to the professional bodies, requirements for vocational and yet linked to vocational knowledge in higher education institutions like Technikons on one hand and linked to the professions on another. The distinction between vocational and professional was not as pronounced. These are the links which seem to be missing in the current reforms. The current standards based approach to the vocational curriculum seems to be using what Young (2005) terms a:

> …functional analysis that was developed by occupational psychologists concerned with job design. The standards-based approach began by identifying and stating curriculum outcomes in terms of what an employee would be expected to do, not what he or she needed to know. Knowledge came second and was only important in so far as it *underpinned* performance (pp.11).

Here, Young is referring to the National Vocational Qualifications (NVQs) of England. He might as well have been referring to the unit standards and vocational / occupational qualifications in the South African NQF.

The new qualifications have increased the supply side of qualifications, whilst the decline of available numbers of qualified artisans is now beginning to be a serious cause for concern for the economic growth and development of the country's infrastructure. This confirms

Young's (2005) observations that in most countries, general education has expanded and employers tend to prefer academic qualifications as criteria for recruiting new employees, and this consequent academic drift has caused a acute shortage of people with higher level technical skills. With the demise of the apprenticeship system, the technical and vocational output in South Africa has been increasingly resembling its counterparts across Africa, a major problem of graduate unemployment, and a growing emphasis for training students for the informal sector and Small Micro and Medium Enterprises (SMMEs) (King & MacGarth, 2002).

So, in the new dispensation, we started off by discarding the 'theoretical subjects' which were seen to be irrelevant, as a way of improving quality. This move has subsequently proved not to have been a wise move as we seem to have attacked the wrong problem. First, Young argues that it is the nature of the knowledge needed to make vocational education to be of acceptable if not higher quality that is a problem here, and that the use of national occupational standards has been an unfortunate pre-occupation in vocational education, which at best was misleading. These remarks are strongly collaborated by our (Umalusi) own research, where we found that the vocational subjects we evaluated are of a lower standards compared to similar general school subjects, and this situation was worse with the new occupational standards (Umalusi, 2005).

Umalusi research shows clearly that while the quality of the old technical and vocational courses have obvious shortcomings, it is extremely difficult to judge the quality of the new 'standards' based qualifications and unit standards. In other words, they are simply empty and devoid of meaning and hence no basis for passing any judgment. Others have noted how we seem to have overloaded the system with unnecessarily detailed internal assessments and quality assurance, without adding any value in the system (Allais, 2003; Young, 2004). Underplaying the power of institutions is also proving to be very problematic. Those who offer the programmes and those who receive the students with vocational qualifications have assumed that once the details have been stated upfront and assessed, this engenders quality. This then raises questions about the nature and form of reforms taken to make education relevant, and also the value of this chosen path in reforming vocational education in SA and elsewhere. This certainly has not improved the perception of users like employers, higher education and the public in general about the status and standards of vocational education.

The new quality assurance mechanisms adopted in the NQF quality assurance model, endorse the need for public criteria in terms of which all institutions and progress could be judged in the form of a quality Management System (QMS). The question that begs an answer is whether this is sufficient for quality improvements. The model adopted by the economic sector quality assurance agencies, relies heavily on the accreditation of centres and monitoring assessment procedures. The model relies on these procedures to make institutions comply. We must heed some of the cautions by Young (2004 (a)) about the new quality assurance system. He notes that that the new system merely shifts power of definitions of quality from the trusted institutions, like schools, colleges and examinations to the hands of a new set of agencies and their associated consultants; does not take into account the problematic nature of the relationship between procedural compliance and quality, and does not remove the links between power and definitions of quality as the new system simply exchanges persons to account to by rules. The South African project of

quality assurance has plunged in knee deep into this model, with very little understanding of what the problems in vocational education are in South Africa. After almost twelve years of this form of reform in vocational education, users are still unsure of the standards of vocational outputs despite all accreditation efforts and the currency or value or the qualifications are still closely tied to institutions; above all the bureaucracy of assessment has expanded to unaffordable levels.

Fortunately, the pendulum seems to be swinging in the other direction now to effect corrective measures in this seemingly chaotic situation we have. The department of education has retreated back to a more centralised systems of curriculum development, with a strong knowledge base, as well as an external examination system, with subject specialists playing a stronger role than stakeholders. This should lead us somewhere in the right direction, but with the multiple quality assurance agencies, many of whom are still trapped in a highly prescriptive and intensely managerialist approach to quality assurance, there are limits to how far we can go. Providers are complying with these requirements because this is the only way they can access training funds available through the Skills Development Strategy of the Department of Labour. Allais (2003) comments that it is possible that an institution may comply with all requirements and still provide poor education. The same observation has been also made by Mokgalane et-al, (2003), about the limited power of an administrative framework to improve quality in education. The pull from the department of education easily also lead to a new language as heightened emphasis on quantitative measures of intuitional performance, (McGrath, 2003). The quantitative indicators which tend to receive increased attention as part of the attempt to analyse and improve institutional performance include number of learners, distribution across programs, teaching staff profile and numbers, pass rates, throughputs rates, cost per learner and many more.

When we remove the notion of vocational education from being purely a labour market tool, or a quantitative commodity we begin to see the problems that continue to bog down vocational education. We actually do not have a clear picture of what the quality issues are in vocational education. It is possible that the agencies and models we have come to accept as establishments, with wisdom about quality, have very little value to add beyond their self serving interests, and at best can act as accountability levers on how institutions spend resources given by governments to them, (Young, 2004). On the question of quality and standards, it looks like in South Africa we have to go back to the drawing board first in order to understand the questions we are asking about vocational education. Unlike in school education quality indicators in vocational education seem to be absent from most systems, in spite of this galloping to the unknown horizon in the name of quality assurance, (McGarth, 2003). In fact, McGarth (2003) is of the opinion that not only in the South African vocational education system, but also internationally, there is a tendency to avoid building a more complex picture, which speaks to qualitative issues in vocational education.

Conclusion

It is argued here that whilst a country like South Africa had to establish a more democratic educational system that places high value on 'quality for all', the country has really not succeeded in charting a coherent path towards achieving 'quality for all'. We started well in tackling issues of equalising provisioning. We succeeded with this project within a short period, but we did not follow through the consequences of our historical deprivation in

major parts of our education system. Young (2004) argues that in South Africa or elsewhere, it is not just the distribution of access that has to be changed but also the criteria of educational success itself, especially in relation to vocational education. In South Africa we do note seem to have necessarily been successful in establishing the quality indicators important across space and time. For example the European Ministers of Education, have reduced their sixteen indicators to only four criteria for measuring quality in the education of the European Union countries (European Commission Directorate for Education and Culture, 2001). We have a vague idea of what these could be in general / school education and seem to be far form reaching consensus in vocational education.

Does quality assurance improve the quality of education? This is the question I set out to examine in this paper. Going by the South African experience, which is not dissimilar to other countries', we may easily conclude that it does not. In fact Allais (2003) firmly believes that there are great limits to what quality assurance can do, conceptually and logistically. She is of the opinion that it can only serve as a check in the system. There is a high degree of truth in this assertion if one goes by the models of quality assurance which run as parallel bureaucracies to the real education system of schools and colleges. Perhaps it is time to question the conceptualisation of quality assurance itself.

Firstly, a useful metaphor for quality assurance is the one that constructs this process as merely a scaffold for building the main theatre. Quality assurance should provide tools to understand quality issues in the system; set the parameters and design for the main building with a full understanding that the real quality issues are to play themselves out in the institutions if not the classroom. If one understands and accepts this metaphor, one appreciates how a scaffold can never be the building, but merely tools to support the building project.

Secondly, quality can be engendered through quality assurance if the starting point is not the procedures, rules and establishment of a bureaucracy to manage quality, but asking the basic question: What are we trying to improve here? or What is the problem? Most quality assurance approaches do not bother to ask that question, instead referenced literature is buzz with models like Total Quality Management (TQM); the Baldridge Awards, the ISOO 9000, etc. (Herstein, 2002; South African qualifications Authority, 2001) These models often have no relation to the real problems in the system. Can we really improve reform and transform education systems if we are unsure of what we need to do away with and what we need to replace? If there is one thing that perhaps quality assurance agencies should strive to do and do well, is to make the education systems in which they operate to understand where the problems are in the system, instead of just galloping to the horizon with their uninformed sense of quality issues. When quality issues are publicly debated, it becomes easier to arrive at some forms of agreements about the indicators which will tell us if and when we have arrived. This is an extremely important element of quality improvement as it sets the agenda in a particular direction, and to go back to our metaphor, sets the shape and size of the building.

There are a number of conclusions which can be drawn from the above discussion. The quality assurance community of a country has a crucial role to play in the improvement of

standards and quality of the education system. First, it can do this by giving the country a meaningful language to talk about quality in its context to which it must attempt to build consensus as a point of reference for measuring quality. Secondly, once the problem is understood, quality assurance communities would assist by stabilising benchmarks which are reasonable to aspire to, as radars for quality improvements. These benchmarks, can then be used as inspection and system assessment tools to gain insights on how the system is performing and reaching these benchmarks. There is nothing about making institutions accountable through accreditation or inspections, but these should be the means of getting institutions to reach and go beyond the set benchmarks, and not the end by itself.

Finally at the heart of quality issues for both general and vocational education, seems to be the matter of curriculum assessment and certification. This is a complex and difficult concept to come into full grips with, partly because in most countries these functions are often separated and also because expertise on vocational curricula is not in abundance. Nevertheless, the quality assurance community, has a huge responsibility to account to the countries citizens and their governments whether or not the learners are being offered curricula of the right quality in general and vocational education, at the right pace, Through the assessment and certification system in place, these agencies are also meeting the accounting responsibility. Countries are competitive and often compare themselves with neighbours and sometimes want to know how they compare to countries, they want to compare with in the global economy villages. (Muller, 2003) is of the opinion that assessment and certification community has the responsibility to let a country know whether it is on course to producing its future citizens. He goes on to warn us on the price a country pays when it does not have these signals or ignores them. To put it more pointedly, he is of the opinion that failing in this regards, results in "the quality of citizens (which) is impaired".

In conclusion, I have started by sketching where South Africa is coming from in relation to quality issues in education. I then later proceed on to sketch the progress which has been made over the years. This, I followed with a description of the different quality assurance interventions made for both the general and vocational education streams and also evaluated the impact these interventions have had in improving quality. My conclusion on the matter is that quality assurance stands a chance to improve the quality of education if a comprehensive and directed intervention is designed and inserted within the quality agenda of education, not as a stand-alone or parallel bureaucracy.

References

- Allais, S. (2003). Outcomes-based qualification driven reform and equivalence in academic, vocational and occupational education. South African Institute of Distance Education, South Africa.
- Allais, S. (2005) What do we mean by standards What do we want from matric? Memorundum, June.
- Bhorat, H. (2005). 'Overview of the Skills Debate in South Africa'. A Paper Presented for the Conference on Higher Education and Development, Reflecting on Challenges, Council for Higher Education in Transformation, Cape Town, South Africa

- Chisholm, L. & Kgobe, P. (2003) Gearing for an Integrated Education System. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers.p.1-26.
- Chisholm, L. (2003) The foundations of Reconstruction and Development of Education and Training. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. P. 87-105.
- Department of Education (2000). Education for All – 2000 Assessment South Africa Report. Pretoria, Department of Education.
- Department of Education (2001). Quality Management Systems in Education. Pretoria, Department of Education.
- Department of Education (2002). Education in South Africa: Achievements since 1994. Pretoria, Department of Education
- Department of Education (2004). An Assessment of 10 years of Education and Training in South Africa. Pretoria, Department of Education.
- Department of Education (2005(a)). A National Framework for Teacher Education. Pretoria, Department of Education.
- European Commission Directorate –General for Education and Culture (2001). European Report on the quality of school education: Sixteen Indicators. Luxembourg: Office for Publications of the European Communities.
- Greenstein, R. & Mkwanazi, Z. (2003) Preparing for Power. In Chisholm, L.; Motala, S. & Vally, S. (eds) South African Education Policy Review. Sandown: Heinemann Publishers. P. 65-85.
- Herstein, J. (2002) Summaries of a collection of articles on Quality Management Systems. Harvard Graduate School of Education.
- Human Science Research Council and the Education Policy Consortium (2005). Emerging Voices: A Report on Education in South African Rural Communities – Research for the Nelson Mandela Foundation. HSRC press, Cape Town. South Africa.
- A King, K and McGarth, S. (2002) Globalisation, Enterprise and Knowledge, Oxford Symposion
- Maassen, P. (1998) Quality Assurance in the Netherlands. New direction for institutional research, no. 99. p.19-28).
- McGarth, S. (2003) Improving quality in South African Further Education and Training Colleges. Unpublished paper for Umalusi Conference. Pretoria: South Africa
- Mail & Guardian (2006). Still no Skills in sight: May 19, 2006.
- Mkwanazi-Twala, Z. Mwiria,K. & Greenstein,R. (2003) The ongoing Process of Change. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. p. 147 -171.
- Motala, S. & Tikly, L. (2003) Towards Transitional Governance. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. P. 29-64.
- Nelson Mandela Foundation (2005) How far have we come? Reflections from the first 10 years of basic Education Transformation. Unpublished paper.
- Mokgalane, E, Vally, S. and Greenstein, R. (2003) Between Vision and Practice: Policy Processes and Impmentation. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. P. 87-105.
- Muller, J. (2003) 'Citizenship and Curriculum: Issues for Assessment and Certification'. Keynote address to the AEAA conference, Cape Town, South Africa. 25-27 August 2003

- Porteus, K. (2003). The Personal and Global: Grinding Gears? Reconsidering Efficiency Debates in general Education. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. p. 745 -800.
- South African Qualifications Act (1995)
- South African Qualifications Authority (2001). Quality Management Systems for Education and Training Providers. Pretoria: South Africa.
- Taylor, N. Muller, J. Vinjevold, P. (2003) *Getting School Working: research and Systemic School Reform in South Africa.* Cape Town: Maskew Miller Longman.
- Tikly, L. & Motala, S. (20030 Changing the System. In Chisholm, L.; Motala, S. & Vally, S. (eds) *South African Education Policy Review*. Sandown: Heinemann Publishers. p.107 -126.
- Trümpelmann, M.H. (1991). The Joint Matriculation Board – seventy five years' achievement in perspective. Cape Town, National Book Printers.
- Umalusi (2004). Investigation into the Senior Certificate Examination: Summary report on the evaluation of the senior certificate examination. Pretoria: South Africa.
- Umalusi (2005). Evaluation of the relative standard of selected NQF level 4 Subjects. Pretoria: South Africa
- Young, M. (2003). Finding our way: Vocational Education in England. Unpublished paper.
- Young, M. (2004(a)). Approaches to quality assurance in the GET and FET bands: Umalusi discussion document. Unpublished paper.
- Young, M. (2004(b)). National Qualifications Frameworks; Their Feasibility and Effective Implementation in Developing Countries. A report prepared for the International Labour Organisation (ILO)