# Developing a Framework for Assessing and Comparing the Cognitive Challenge of Home Language Examinations



**UMALUSI**

Council for Quality Assurance in
General and Further Education and Training

10 YEARS
SETTING STANDARDS IN EDUCATION AND TRAINING

# Developing a Framework for Assessing and Comparing the Cognitive Challenge of Home Language Examinations

**Cheryl Reeves**

July 2012

# Acknowledgements

# Acronyms and Abbreviations

| | |
|---|---|
| ASC | Assessment Standards Committee |
| CASS | Continuous Assessment |
| DoBE | Department of Basic Education |
| DoE | Department of Education |
| FET | Further Education and Training |
| GEFT | General and Further Education and Training |
| HG | Higher Grade |
| HL | Home language |
| IRT | Item Response Theory |
| NATED | National Education (curriculum 550) |
| NCS | National Curriculum Statements |
| NSC | National Senior Certificate |
| OBE | Outcomes-based education |
| PIRLS | Progress in International Reading Literacy Study |
| SG | Standard Grade |

# Contents

# Contents continued

# List of tables

# List of tables

# Executive Summary: Part 1

The focus of this research report is on developing a framework for Umalusi for assessing and comparing the **cognitive challenge** of the National Senior Certificate (NSC) **Home Language examinations**.

## Research question
The main research questions for the report are:

> **What framework could best serve for developing a 'cognitive challenge' instrument for assessing and comparing the National Senior Certificate (NSC) Home Language examinations in particular, as well as Home Language examinations for similar qualifications?**
>
> **What are the key criteria for instrument(s) for evaluating and comparing the standards of Grade 12 Home Language examinations?**

## Origin and background to the report
Umalusi has identified the need to evaluate and compare the standards of Grade 12 Home Language examinations across all eleven official South African languages.

Over the past few years, there has been growing concern among members of Umalusi's Assessment Standards Committee (ASC) about the results of some of the indigenous African languages' Home Language examinations.

In particular, there is concern about high pass rates or high percentages for minority languages such as isiNdebele, SiSwati, Xitsonga and Tshivenda. Unlike other official languages, almost all candidates writing Home Language examinations in these languages appear to be passing or achieving higher pass rates.

In the current system, unlike the old NATED 550 system, a single set of Home Language examination papers are expected to discriminate among extremely high-achieving students (who would have got an A-grade on the old NATED 550 Higher Grade papers), students performing at low levels (who would have passed on the old Standard Grade papers), and students performing at levels in between.

Thus it is very important that Grade 12 examinations cover a sufficiently wide range of questions at different levels of cognitive challenge. For a range of cognitive challenge to be evident in examination papers, there needs to be appropriate coverage of a variety of types of cognitive demands (lower-, medium- and higher-level cognitive demands).

Umalusi has developed an instrument specifically for analysing whether the types of cognitive demand of the Grade 12 **Home Language** examinations are comparable. The taxonomy of cognitive demands used in Umalusi's Home Language examination paper analysis instrument is based on **the Revised Bloom's Taxonomy**, and has five categories of cognitive demand, namely: (1) Conceptual knowledge, (2) Comprehension (Understanding); (3) Application; (4) Analysis and Problem-solving; (5) Evaluation and Synthesis (Creation).

However, a key challenge has arisen with respect to Umalusi's Home Language instrument.

In 2010 and 2011, the Department of Basic Education's Home Languages examiners and moderators used a different taxonomy, **Barrett's Taxonomy**, for setting and moderating Home Language examination papers. The specific purpose of this taxonomy is to assess questions that measure reading comprehension.

Because a different taxonomy was used for setting and moderating the Home Language examination papers from the taxonomy Umalusi used for evaluating the cognitive demand of examination papers, Umalusi was unable to make a decision regarding compliance of the examinations in terms of the relative proportions of questions at particular cognitive levels as specified in the Grade 12 *Examination Guidelines* reflected in Table A below.

**Table A: Home Language Examination Guidelines requirements in terms of Cognitive levels**

| Cognitive levels based on Barrett's Taxonomy | Proportions of marks |
|---|---|
| 1. Literal Comprehension | Lower level |
| 2. Reorganization | 40% of total marks |
| 3. Inferential Comprehension | Medium level<br>40% of total marks |
| 4. Evaluation | Higher level |
| 5. Appreciation | 20% of total marks |

No simple one-on-one correspondence could be established between the categories and levels in Umalusi's instrument (based on the Revised Bloom's Taxonomy), and the Barrett Taxonomy used by the Department of Basic Education.

This situation highlighted the need for Umalusi to review and reconsider the use and application value of both taxonomies of cognitive demand and Umalusi's Home Language 'cognitive challenge' instrument.

### Purpose and aim of this report

This report evaluates different taxonomies of cognitive demand used for categorizing the cognitive challenge of examination questions, in terms of what they offer, and where they have limitations. It also considers ways in which Umalusi's current 'cognitive challenge' framework could be made more suitable for assessing Home Language examinations.

### Research approach

The research approach involved:

- an examination of the instrument based on Barrett's Taxonomy used by the Department of Basic Education to set Home Language examinations, and the instruments used by Umalusi in its comparative analysis of the 2008-2010 Home Language examination papers.

- an examination of the *Subject Assessment Guidelines* (DoE, 2008) for assessment of Home Languages in Grades 10-12; the *Guidelines* for the setting of Grade 12 examinations in Languages (Home Languages): Papers 1-3 (DoE, 2009a, b, c); 2008 to 2010 Home Language examination papers (Papers 1 – 3); and the associated memoranda for each of the selected languages.

- a literature survey on the Revised Bloom's Taxonomy versus Barrett's Taxonomy.

- participation in the training of Umalusi's 2011 teams of Home Language examination evaluators for the eleven official languages so as to incorporate a more empirical

approach through observation of the application value of the taxonomy used in actual examination papers.

- reading Umalusi's draft report on the *Comparative analysis of the National Senior Certificate Home Language examinations, 2008-2010* (Umalusi, 2011).

- synthesis of available evidence from the literature review together with data and ideas generated from the documents and meetings.

- drafting a 'cognitive challenge' framework and instrument(s) for assessing Home Language examination papers.

- presentation of the draft 'cognitive challenge' framework and draft instrument(s) to Umalusi's Research Forum.

- using the discussion with the Research Forum to finalise the 'cognitive challenge' framework and instrument(s) for assessing Home Language examination papers.


## Main findings

This section summarises the main findings with regard to

1. Taxonomies of cognitive demand of examination questions;

2. Degree of difficulty of examination questions; and

3. Other factors that have a material influence on the cognitive challenge of Home Language examination papers.

---

**Summary Main Findings: Taxonomies of Cognitive Demand**

1. Any **cognitive demand** taxonomy selected or constructed for evaluating Home Language examinations papers **needs to cover all aspects of the examination papers**, for example, literature, grammar, creative writing, etc.

2. The **Revised Bloom's Taxonomy** (used for Umalusi's Home Language instrument) **does not include cognitive processes specifically involved in comprehending written text. Barrett's Taxonomy is not designed to evaluate questions that test grammatical skills in isolation from the text** (for example, grammatical rules), **or tasks which require students to generate or create their own text** (for example, a full composition, or a letter), or other original product (such as designing an advertisement).

3. The **Revised Bloom's and Barrett's taxonomies could be combined to complement the weaknesses and strengths of each other.** Ultimately, what is important, is that evaluators are able to (1) discriminate questions more broadly as requiring lower, medium, and higher level cognitive processes; (2) check that an appropriate range of cognitive demands are made on students; and (3) ensure that higher level cognitive demands are not under-represented in examination papers.

4. **The assumption cannot be made that 'alignment' with the Department of Basic Education's allocated proportions of questions** at lower, medium and higher order cognitive demand levels means that 40% of the examination questions are relatively easy for the average Grade 12 student to answer; 40 % of the questions are moderately challenging for the average Grade 12 student to answer; and 20% of the questions are difficult for the average Grade 12 student to answer, and **allows for A-grade students to be discriminated from students performing at lower levels.**

---

**Summary Main Findings: Degree of Difficulty of Examination Questions**

1. It seems that **cognitive levels on their own do not necessarily distinguish between degrees of difficulty of questions**. Questions that align to a particular 'type of cognitive demand' are not always 'easier' than other questions that align to the same cognitive demand level. For example, a 'recall' question can ask students to restate or explain a simple fact, or a much more abstract theory or complex content, the latter being much more difficult to accomplish.

2. Umalusi's Home Language instrument has attempted to address this challenge by offering evaluators wider options in terms of 'Easy', 'Moderate' or 'Difficult' degrees of difficulty for each category of cognitive demand'. **Umalusi's modifications** have overcome some of the limitations, but they have also created new challenges. There is a danger that the finer categorization **'tick box' options** of three different difficulty levels for each cognitive process category **are used in a mechanical manner making judgments overly procedural, thereby constraining the 'expert' judgment of evaluators**.

3. There is a need to **take into account changes in cognitive demand expectations across different grade levels, and the difficulty of examination questions specifically for Grade 12 level students**. For example, the same question requiring inferential reasoning can be far more difficult for a Grade 6 student than for a Grade 12 student. What needs to be taken into account, is the conceptual demand that a particular question makes on the cognitive schema of a typical Grade 12 student.

4. To adequately capture the complexity of this judgment task, it may be necessary **to combine relatively low-inference 'tick box' ratings of 'type of cognitive demand'** for specific questions **with more implicit expert judgments of the difficulty of examination questions and texts for the average Grade 12 student**. What is difficult to specify, is what makes a particular 'evaluation' question (for example) difficult for the average Grade 12 student, when the same question may be easy for a (Home) Language university graduate.

---

**Summary Main Findings: Other Factors that have a Material Influence on the Cognitive Challenge of Home Language Examination Papers**

1. **The level of difficulty of the (written or graphic) texts that have to be read also influence the level of cognitive challenge** – regardless of the type of questions asked about the texts. For example, literal comprehension of a simple contemporary magazine article is easier that literal comprehension of a classical work such as one of Shakespeare's dramas (in terms of content, vocabulary, sentence and organisational structure, register, diction, literary techniques, abstractness of ideas and imagery, and background knowledge required).

2. **Texts selected may differ significantly across different years, and across different Home Language examinations, in terms of their intrinsic difficulty, making the examination more (or less) cognitively challenging.** Differences in the degree of challenge in prescribed set works across Home Languages can make an immense difference to the cognitive challenge of the different Home Language Literature examinations.

3. **A cognitive challenge framework for Home Language examinations also needs to take into account the level of writing demands that questions make on students.** For example, a short answer requiring writing one word answer, a phrase, or a simple sentence is easier to write than responses that require more complex sentences, a paragraph or a full essay or composition. A question that requires low levels of cognitive demand such as recalling and reconstructing an ordered sequence of events, for example, could entail writing a few sentences, or a whole essay.

**Recommendations: A framework for assessing and comparing the Cognitive Challenge of Home Language Examinations**

The report makes **FOUR** main recommendations.

It recommends that Umalusi:

1. **Re-introduces simple categories of judgment for 'type of cognitive demand', which combine the strengths of Barrett's and the Revised Bloom's Taxonomies. The categories should be analysed as low, medium and high levels of cognitive demand.**

2. **Includes a more inferential expert judgment of the difficulty level of each question for Grade 12 students. Umalusi's evaluators are not novices in their respective fields. Along with checklist categories of cognitive demand, they should be also be given opportunities to use their knowledge of the Home Language, and experience of teaching it, to exercise their expert judgment about the difficulty levels of examination questions and texts for Grade 12 students.[1]**

3. **Takes into account the writing demands of responses to examination questions (for example, whether responses require writing a one word answer, a phrase, a simple sentence, more complex sentences, a paragraph, or a full essay or composition).**

4. **Incorporates judgment of the degree of challenge of the written or graphic texts that students at the Grade 12 level are required to read (prescribed texts, reading passages or other source material) for Home Language examinations.**

The report for Part 2 of the research provides

a) a framework for judging levels of difficulty of examination questions which evaluators could use to make tacit expert notions of question difficulty more explicit; and

b) a new instrument for Umalusi's comparative analysis of Home Language examinations.

---

[1]Evaluator's views about the difficulty levels of each question for Grade 12 students can be compared with item analyses after students have written the examination.

# Executive Summary: Part 2

## Introduction

This report forms the second part of Umalusi's research into improving its framework for assessing and comparing the standards of the National Senior Certificate (NSC) in the different Home Languages across different years. The focus in Part 1 of the research was on the use and application value of taxonomies of cognitive demand for analysing the cognitive challenge of NSC Home Language examination questions.

The specific focus of Part 2 of the research is on approaches to describing and classifying varying degrees of difficulty of NSC Home Language examination questions prior to examination writing.

## Research question

The main research questions for Part 2 of the report are:

> **What framework could best serve as guidelines for assessing and comparing the difficulty level of questions in National Senior Certificate (NSC) examinations?**
>
> **What are key categories and concepts for consideration when judging question difficulty?**

## Origin and background to the report

One of the main tools recruited in Umalusi's instruments for distinguishing the levels of cognitive challenge of examination questions has been a taxonomy of cognitive demand based on the Revised Bloom's Taxonomy (Anderson and Krathwohl 2001; Anderson 2005). Panels of evaluators comprising experienced teachers, subject experts, and higher education experts, have used Umalusi's taxonomy to rate the cognitive demand levels of individual examinations questions so as to make judgements about the relative standards of different examination papers across a number of years.

The purpose of the report, *The challenge of cognitive demand (Part 1 of Developing a framework for assessing and comparing the cognitive challenge of Home Language examinations)*, was to evaluate the various taxonomies of cognitive demand used for categorising Language examination questions, in terms of what they offered, and where they had limitations. The research confirmed the importance of ensuring that a range of cognitive demands are made on students in Grade 12 examinations. Findings endorsed the need to check that examination papers are not dominated by questions that require reproduction of basic information or replication of basic procedures.

A key recommendation was that, for this purpose, the Revised Bloom's Taxonomy (used by Umalusi) and Barrett's Taxonomy (used by the Department of Education) be combined to complement the weaknesses and strengths of each other. The report provided a taxonomy that 'married' both taxonomies.

However the investigation into cognitive demand taxonomies also showed that there are factors other than the level of cognitive demand that influence the degree of challenge of an examination paper. For example, individual questions cannot be evaluated in isolation from other aspects of the item as a whole such as the information or 'texts' (passages, diagrams, tables, graphs, pictures, cartoons, etc.) that accompany each question, or the written or graphic texts that students are required to produce in order to respond. The report pointed to differences in the degree of challenge in prescribed set works that can make an

immense difference to the difficulty of the different Home Language Literature examinations and the demands of questions.

It drew attention to the fact that a particular 'evaluation' question (for example) may be more difficult for a typical Grade 12 student than for a university graduate. A question requiring inferential reasoning may be far more difficult for the average Grade 6 student than for the average Grade 12 student. As questions must be of appropriate difficulty for the students to whom they are administered, difficulty is to a certain extent, relative to the particular reference group.

The report suggested that in addition to relatively low-inference ratings of 'type of cognitive demand' for specific questions, evaluators should also exercise expert judgements about whether each examination question is 'Easy', 'Moderate' or 'Difficult' for the 'average' Grade 12 student. It argued that expert judgment of levels of difficulty of examination questions or items usually involves using multiple and inter-dependent criteria.

The complexity of the judgement task lies in the evaluators' ability to recognise subtle interactions and links between different aspects of each question's difficulty and make holistic judgements. Such judgements are complicated, non-linear processes.

A rigid judgement strategy for assessing the difficulty of examination questions involving the use of prescribed criteria runs the risk of evaluators losing a sense of judgement as a coherent whole. Using 'tick box' options for grading the difficulty level of examination questions or items tends to result in procedural compliance taking precedence over expert judgement.

The process of trying to match questions to pre-set definitions of levels of difficulty distracts evaluators from the idea that, what they are being asked to do, is to use their internalised expertise and experience to determine whether the range of questions in examinations make it possible to discriminate between high-achieving students, students performing at low levels, and students performing at levels in between.

A key recommendation of Part 1 of the report is that evaluators be given opportunities to use their knowledge of the subject, their experience of teaching it, and/or of marking Grade 12 examination scripts, to exercise a more nuanced and inferential judgements of question difficulty. Paradoxically, if Umalusi's evaluation process is to gain public confidence, the judgement process needs to be transparent. What makes the process more or less transparent is the extent to which what makes one question more difficult than another for candidates is explicitly addressed.

Evaluators need to provide an account of how their decisions are made and to justify their decisions. For this purpose, they need to able to identify where the source of difficulty or ease in each question resides for Grade 12 candidates. However, the tacit nature of the expert knowledge employed makes it difficult for them to articulate their understandings or generate a common understanding of what constitutes levels of difficulty (Fisher-Hoch and Hughes, 1996).

The conclusions and recommendations of Part 1 of the Home Language report thus raised the need for an investigation into guidelines for assessing the difficulty level of examination questions and identifying a judgement strategy that renders the expert judgements process transparent.

## Purpose and aim of this report

The aim of this report is to

- identify an approach to assessing the difficulty level of examination questions prior to examination writing that goes beyond a 'tick box' approach;

- provide a framework that could be used by evaluators when judging levels of difficulty of examination questions and which evaluators could use to make tacit notions of question or item difficulty or 'easiness' more explicit and transparent; and

- present a new instrument for Umalusi's comparative analysis of Home Language examinations.

## Research approach

The research approach involved:

- Surveying literature on standard setting processes and item or question difficulty. The main objective was to review approaches to describing and classifying varying degrees of difficulty of examination questions and identify possible guidelines for making expert judgement of examination questions.

- Using the above to draft a new version of Umulusi's instrument '*A comparative analysis of the National Senior Certificate Home Language examination papers*'.

- Participation in the piloting of the revised the instrument with Umalusi's 2012 teams leaders of Home Language examination evaluators for the eleven official languages in June 2012, and in a workshop (2-4 July 2012) designed to provide a platform for all Home Language examination role players to discuss and critique the proposed conceptual framework for the new instrument.

- Using the June pilot and the July workshop to finalise the instrument for comparing the 2009-2012 Home Language examination papers.

- Presentation of the research findings and the revised instrument to Umalusi's Research Forum.

- Using the discussion with the Umalusi' Research Forum to finalise the report and the instrument.

## Main findings and recommendations

This section summarises the main findings and recommendations with regard to

- the framework that evaluators should use for judging the difficulty level of questions in National Senior Certificate (NSC) Home Language examinations; and

- the new instrument for Umalusi's comparative analysis of Home Language examinations.

**Summary main finding: A framework for thinking about question difficulty**

The investigation found that:

1.  What is needed, is a framework which helps to make explicit what is tacit, but which is not written in a prescriptive way that prevents evaluators from grappling with nuances and making connections.
2.  The most realistic approach appears to be for evaluators to have a simple frame of reference for thinking about item or question difficulty as a means of guiding and supporting their judgement of levels of difficulty.
3.  Ideally a framework for assessing the difficulty level of examination questions need to act as a heuristic device which helps to shape and guide rather than prescribe, or dictate and control, the judgement process.
4.  The framework should serve to influence evaluators' perceptions of question difficulty and help them develop a more explicit understanding of what aspects make a question difficult or easy.

---

**Recommendation: A framework for thinking about question difficulty**

The four-category conceptual framework for thinking about item or question difficulty provided by Leong (2006) of the *Singapore Examinations and Assessment Board* was identified as most appropriate and useful for this purpose.

The categories for consideration when thinking about the level of difficulty of examination questions for Grade 12 students of difficulty are:

*   **Content difficulty**, which refers to the difficulty in the subject matter and/or concept/s assessed or involved. In this judgment, difficulty lies in the academic and conceptual demands that questions make and the grade level boundaries of the various 'elements' of the knowledge domain (facts, concepts, principles and procedures associated with the subject).

*   **Stimulus difficulty**, which refers to the difficulty of the linguistic features of the question and the challenge that candidates face when they attempt to read and understand both the words and phrases in the question, and the information or 'texts' (diagrams, tables and graphs, pictures, cartoons, passages, etc.) that accompany the question.

*   **Task difficulty**, which refers to the difficulty that candidates confront when they try to generate or formulate an answer.

*   **Expected response difficulty**, which refers to difficulty imposed by examiners in a mark scheme and memorandum. Mark allocations affect the amount and level of answers students are expected to write.

The framework provides a range of examples of aspects of each category. The underlying assumption is that judgement of question difficulty is influenced by the interaction and overlap of a variety of aspects in the four different categories.

For each general category, Leong (2006) draws a further distinction between 'valid' and 'invalid' sources of question difficulty or easiness. 'Valid difficulty' or 'easiness' has its source in the requirements of the question and is *intended* by the examiner (Ahmed and Pollit, 1999).

Invalid sources of difficulty or easiness are defined as those features of question difficulty or easiness that were *not intended* by the examiner. Invalid sources prevent the question from assessing what the examiner intended and are likely to prevent candidates from demonstrating their true ability or competence (Ahmed and Pollit, 1999, see also Hannah Fisher-Hoch and Hughes, 1996:2). They are factors irrelevant or indirect to the construct being measured. The framework includes examples of probable invalid sources of difficulty for each of the four categories of difficulty.

**Recommendation: The new Home Language instrument**

The new instrument for Umalusi's comparative analysis of Home Language examinations can be found in **Appendix A** of this report. The following features of the instrument should be noted:

1. The new instrument draws **a distinction between judging the level of cognitive demand of examination questions and judging the difficulty level of examination questions**.

2. Evaluators need to identify **questions deemed** to be 'easy', 'moderately challenging', 'difficult' and **'very difficult' for the average Grade 12 student to answer. The forth level, 'very difficult'**, has been included in the levels of difficulty of examination questions to enable evaluators to identify questions where the skills and knowledge required to answer them **allow for A-grade students** (extremely high-achieving/ability students) **to be discriminated from other high ability students**. The aim is to ensure that there are sufficient questions that discriminate well amongst higher ability candidates.

3. **Evaluation team members** are required to use the difficulty framework provided to identify and list the main sources of difficulty or 'easiness' (Content, Stimulus, Task and/or Expected Response) in each question, and **provide** their **reasoning for the level of difficulty selected** for each examination question.

4. Home Language evaluation team members first use the framework to make **decisions** about question difficulty **on an individual basis** and, only after this, **reach a consensus rating** through discussion with other members of their language teams.

5. In addition to judging the difficulty level of each examination question, evaluators are also required to **judge the difficulty of the source or stimulus material** (reading selection or passage, visual text or source material including prescribed novels, poems, and dramas) that students are required to read or refer to in each of the Home Language examination papers. They are asked to use ratings provided in the instrument to evaluate whether the stimulus or source material in each of the Home Language examination papers makes low, medium, or high 'reading' demands on the average Grade 12 examination candidate.

# 1. THE 'CHALLENGE' OF COGNITIVE DEMAND

## Chapter 1: Context, Rationale, Scope and Aim

## 1.1 Introduction and objective

The focus of this research report is on developing a framework for Umalusi for assessing and comparing the cognitive challenge of the National Senior Certificate (NSC) Home Language examinations.

Chapter 1 provides

- the context and rationale for the study;

- scope and purpose of the research;

- timeframe;

- research approach used; and

- aim of this report.

It concludes with an outline of the structure of the rest of the report and its chapters.

## 1.2 Context and rationale

### 1.2.1 Monitoring standards

Umalusi is mandated to set and monitor educational standards in General and Further Education and Training (GEFT) in South Africa. Standards in education systems have to do with 'cognitive challenge', or degree of difficulty, and are traditionally revealed in the curriculum and examinations. The level of difficulty in examination systems is revealed most clearly through the examination questions.

Since 2003, Umalusi has conducted several research studies that have investigated examination standards. For example in 2008, Umalusi conducted research on the National Senior Certificate (NSC) examinations, commonly known as 'Matriculation' or Grade 12, in order to gain an understanding of the standards of the new examinations (first introduced in 2008) relative to those of the previous NATED 550 examinations.

The purpose of this *Maintaining Standards* research (Umalusi, 2009a, b and c; 2010) was to ensure continuity of standards between the old qualification and the new qualification. It was thought that a comparison of the final school exit examinations would provide an indication of whether Grade 12 students were being required to perform at similar levels in the old and new examination systems. Previous Umalusi research (Umalusi, 2004) had focused on the standard of the Grade 12 examinations over the period 1992 – 2003.

## 1.2.2 Tools for making judgments

The above research revealed the need to develop tools with which judgments could be made about the standards of examinations. In particular, it raised the question of how *comparative* judgments can be made about the difficulty of examined curricula. The research created the need to impose some sort of standardized criteria to compare the degree of difficulty of examinations associated with the different curricula for subjects and qualifications, in relation to each other (for example, the comparative difficulty of South Africa's school exit qualifications in relation to other, similar international qualifications), and over a number of school years. To ensure the integrity of results, Umalusi needed to have an informed understanding of the cognitive challenge of the examinations.

The research created the need to develop a comprehensive and systematic tool for

a)  determining the degree to which the respective examinations are consistent with difficulty levels in their own curriculum statements;

b)  comparing difficulty levels of the same high stakes examinations over a number of years so as to ensure that successive examinations are equally challenging, and that consistent standards are being maintained over time.

c).  evaluating the difficulty levels of high stakes subjects' examinations in relation to other similar high stakes subject examinations (for example, for different qualifications), to establish whether examinations are equally challenging.

Umalusi, with the assistance of various subject, curriculum and assessment experts, developed a tool to evaluate examinations for various National Senior Certificate (NSC) subjects. After considerable debate, the Revised Bloom's Taxonomy (Anderson and Krathwohl, 2001; Anderson 2005) was recruited as the basis for developing an instrument for evaluating the levels cognitive demand of examination papers.

This taxonomy builds on Bloom's cognitive domain Taxonomy, one of the best known educational taxonomies for assessing cognitive complexity. Most of the National Senior Certificate subject examination specifications state that examination papers should be set in such a way that they reflect proportions of questions at various levels of cognitive demand based on Bloom's Taxonomy. It made sense for the Department of Education to refer to Bloom's Taxonomy as most teachers and educational experts are familiar with it.

Umalusi wanted to establish whether it is possible to have a single 'generic' tool that could be utilized to make judgments about levels of cognitive demand of examinations throughout all subject areas. It was felt that using the same categorical distinctions or cognitive demand levels across different subjects would make it easier to make comparisons between subjects. The Revised Bloom's Taxonomy appeared to provide a suitable framework for measuring levels of complexity of cognitive operations in examination papers in different subjects.

The cognitive processes in the Revised Bloom's Taxonomy closely resemble those found in Bloom's Taxonomy. Like the original taxonomy, the Revised Taxonomy identifies levels of cognitive processes in a hierarchy, from the highest to the lowest level of complexity. The Revised Bloom's Taxonomy was selected for use by Umalusi, because it also separates forms of knowledge (four types) from types of cognitive process (six levels).

The viability of a single instrument approach was tested empirically when the Umalusi instrument was used to analyse and evaluate actual examination papers.[2] When evaluators applied and used the instrument to evaluate question papers for the various subjects, they found that they needed to adapt and refine the 'generic' instrument Umalusi had provided, usually simplifying it but, more importantly, making it more appropriate for their particular subject. In other words, there was a strong tendency amongst the various teams to move from a single set of cognitive process categories to multiple instruments with different sets of categories for describing cognitive challenges within different subjects.

Over time, additional grids were developed for the various subjects, which attempt to distinguish between types of cognitive processes as well as the degree of difficulty of cognitive processes (see Chapter 2 for details).

## 1.2.3 Evaluating standards of Home Language examinations

More recently, Umalusi identified the need to evaluate and compare the standards of Home Language examinations across all eleven official South African languages. Over the past few years, there has been growing concern among members of Umalusi's Assessment Standards Committee (ASC) about the results of some of the indigenous African languages' Home Language examinations. In particular, there is concern about high pass rates and/or high percentages for minority languages such as isiNdebele, SiSwati, Xitsonga and Tshivenda. Unlike other official languages, almost all candidates writing Home Language examinations in these languages are passing or attaining higher than average percentages.

In 2010, the Home Language instrument that Umalusi had developed was used to determine whether the types of cognitive demand and degrees of difficulty of the Grade 12 Home Language examinations for six of the eleven official languages (Afrikaans; English; IsiNdebele; IsiXhosa; IsiZulu; and SiSwati) are comparable over a three year period (2008 – 2010), and across the six languages (for example, whether higher level cognitive demands are equally represented in question papers) (Umalusi, 2011). However, a key challenge, or limitation, arose with respect to Umalusi's Home Language instrument.

This challenge is that, whereas the Department of Basic Education's (DoBE) examiners in other subjects draw on Bloom's Taxonomy in setting examination papers, the Department's Home Languages' examiners and moderators used a completely different taxonomy, Barrett's Taxonomy, for setting and moderating at least two of the three Home Language examination papers in 2010 and 2011.[3]

For the Home Language examinations, three papers are set externally. These are:

- Paper 1: Language in Context;
- Paper 2: Literature; and
- Paper 3: (Creative) Writing.

---

[2]Judgments about examinations were also guided by other criteria such as content coverage, constructs employed, variety of task types, length of paper, language and cultural bias, clarity of instructions, organization of paper, additional criteria supplied to candidates, and relationship between the paper and marking memorandum.

---

[3]The *Guidelines for the Setting of Grade 12 Examinations in Languages: Paper 1, Home Language, First Additional Language, and Second Additional Language* (DoE, 2009b) do not explicitly state which taxonomy should be used. Information from marking panels suggests that, in 2009, Bloom's Taxonomy was used for setting Paper 1. The *Examination Guidelines Grade 12, Literature: Paper 2* (DoE, 2009a) states that Barrett's Taxonomy was used for setting both Paper 1 and Paper 2.

Paper 4 takes the form school-based assessment of oral tasks. According to the *Examination Guidelines Grade 12, Literature: Paper 2 Home Language and First Additional Language, and Paper 1: Section D: Second Additional Language* (Department of Education, 2009a: 23), examination papers should be set in such a way that they reflect the following proportion of questions at the various levels of cognitive demand, based on Barrett's Taxonomy.

**Table 1.1: Home Language Examination Guidelines requirements in terms of Cognitive Levels**

| Cognitive levels | Proportions of marks |
|---|---|
| 1. Literal Comprehension | 40% of total marks |
| 2. Reorganization | |
| 3. Inferential Comprehension | 40% of total marks |
| 4. Evaluation | 20% of total marks |
| 5. Appreciation | |

The teams of Home Language examination evaluators[4] for Umalusi's 2010 report on the standards of the Afrikaans, English, isiNdebele, isiXhosa, isiZulu, and SiSwati Home Language examinations administered to South African Grade 12 students used the instrument that Umalusi had developed for assessing the levels of cognitive demand and degree of difficulty of Home Language examinations. But, because such a different taxonomy was used for setting the Home Language examination papers from the taxonomy Umalusi used for evaluating the examination papers, Umalusi was unable to draw conclusive facts about compliance of the examinations in terms of the weighting of the cognitive categories specified in the *Examination Guidelines Grade 12 Literature: Paper 2, Home Language and First Additional Language, and Paper 1: Section D: Second Additional Language* (DoE, 2009a).

No simple one-on-one correspondence could be established between the categories and levels in Umalusi's instrument based on the Revised Bloom's Taxonomy and Barrett's Taxonomy. Consequently, Umalusi was unable to ascertain the degree to which the respective Home Language examination question papers adhere to the national assessment policy.

In the previous system, subjects for the Grade 12 school qualification (NATED 550 curriculum) were generally offered on two main levels: Higher Grade (HG) which was supposed to be more cognitively challenging, and Standard Grade (SG) which was designed as an easier alternative. Where the Higher Grade papers in the previous system discriminated students achieving at the highest levels, the 'high-level questions' in the National Senior Certificate (NSC) papers now need to accomplish the same task.

Thus the relative proportions of examination questions at particular cognitive and difficulty levels are very important in the current system, where a single set of papers is expected to discriminate between extremely high-achieving students (who would have got an A-grade on the old Higher Grade), students performing at low levels (who would have passed on the old Standard Grade), and students performing at levels in between.

---

[4]These teams of four people normally comprise a university subject methodology specialist, with at least five years' experience; a subject advisor who has no less than five years' experience; and teachers with at least five years Grade 12 teaching experience.

For this purpose, there needs to be an appropriate range of cognitive challenges demanded of students across the examination papers, and the proportions of cognitive demands should match those stipulated in the *Subject Assessment Guidelines: English, Home Language, First Additional Language, Second Additional Language (DoE, 2008a), or in the Grade 12 Home Language Examination Guidelines* (DoE, 2009a, b, c).

This situation highlighted the need for Umalusi to review and reconsider the use and application value of the two taxonomies: the taxonomy used in Umalusi's Home Language instrument based on the Revised Bloom's Taxonomy and Barrett's Taxonomy, which forms the basis of the instrument used by the Department of Basic Education.

## 1.3  Scope and purpose

The purpose of this report is to consider a cognitive challenge framework for assessing the Home Languages as part of an examination indicator system to characterize and describe the cognitive demands made in the Home Language examination papers.

For Umalusi's purposes the central questions that the framework needs to help determine are: whether the cognitive challenge of the skills and knowledge examined across the eleven official Home Languages are consistent with each other and in relation to previous years; and if not, how they differ.

Clearly what is needed for this process is:

• A definition of what counts as evidence of 'cognitive challenge'.

• Sufficiently detailed language for measuring cognitive challenge to meet Umalusi's and the Department of Basic Education's needs.

• Criteria which allow individual questions to be rated as proof of the degree of cognitive challenge.

• Assessment of the cognitive challenge of examination questions that goes beyond a tick box approach, and assesses whether the range of cognitive challenge in examination paper questions makes is possible to discriminate between extremely high achieving students, students performing at low levels, and students performing at levels in between.

## 1.4  Research approach

The research approach for Part 1 takes the form of

• an examination of instrument based on Barrett's Taxonomy used by the Department of Basic Education to set Home Language examinations, and the instruments used by Umalusi in its comparative analysis of the 2008 – 2010 Home Language papers;

• an examination of the *Subject Assessment Guidelines* (DoE, 2008) for assessment of Home Languages in Grades 10-12; the *Guidelines* for the setting of Grade 12 examinations in Languages (Home Languages): Papers 1 – 3 (DoE, 2009a, b, c); 2008 to 2010 Home Language exam papers (Papers 1 – 3); and the associated memoranda for each of the selected languages;

• a literature survey on the Revised Bloom's Taxonomy versus Barrett's Taxonomy. The main objective of the literature review is to provide an underlying conceptual framework for comparing the levels of cognitive challenge of the Home Language examination papers;

• participation in the training of Umalusi's 2011 teams of Home Language examination

evaluators for the eleven official languages (including Sepedi, Setswana, Sesotho [Sotho languages]; Tshivenda and Xitsonga}. The main purpose of participating in the training of Umalusi's 2011 teams of examination evaluators was to incorporate a more empirical dimension or approach into the study through observation of the application value of the taxonomy used in actual examination papers.

- reading Umalusi's draft report on the *Comparative analysis of the National Senior Certificate Home Language examinations, 2008 – 2010* (Umalusi, 2011). Evaluation teams were asked to provide feedback on the usefulness of Umalusi's instrument. The reports provide their feedback on whether or not the compilers think that using the instrument helped to provide reliable indicators of the cognitive challenge of the Home Language examinations.

- synthesis of available evidence from the literature review together with data and ideas generated from the documents and meetings to answer the following key research question:

   **What framework could best serve for developing a 'cognitive challenge' instrument for assessing and comparing the National Senior Certificate (NSC) Home Languages examinations in particular, as well as Home Language examinations for similar qualifications? What are key criteria for instrument(s)?**

- drafting a 'cognitive challenge' framework and instrument(s) for assessing Home Languages examination papers.

- presentation of the draft 'cognitive challenge' framework and draft instrument(s) to Umalusi's Research Forum meeting held on 5 October 2011.

- using the discussion with the Research Forum to finalise the 'cognitive challenge' framework and instrument(s) for assessing Home Languages examination papers.


# 1.5  Aim

The aim of this report is to

- try to establish the optimal application of a taxonomy and instrument(s) to make assessment of the cognitive challenge of questions in Home Language examination papers more reliable (Part 1);

- identify different pre-examination approaches to describing and classifying difficulty levels of examination questions (Part 2);

- compare different methods and models for judging question difficulty for candidates (Part 2);

- propose guidelines that could be used by evaluators when judging levels of difficulty of examination questions and which evaluators could use to make tacit notions of question or item difficulty or 'easiness' explicit or public (Part 2); and

- present a new instrument for Umalusi's comparative analysis of Home Language examinations (see Appendix A).

# 1.6  Outline of report

**Chapter One** frames the report by providing the rationale and background to the study and the research approach used in Part 1.

**Chapter Two** provides an overview and comparison of three question taxonomies - Bloom's Taxonomy, the Revised Bloom's Taxonomy, and Umalusi's modified version of the Revised Bloom's Taxonomy. It discusses key debates and concerns relating to each of the taxonomies, and their use and application in educational and examination contexts.

**Chapter Three** presents Barrett's Taxonomy. As examples of other taxonomies and instruments based largely on Barrett's Taxonomy, the chapter outlines the Progress in International Reading Literacy Study (PIRLS) Taxonomy, and the modified version of Umalusi's instrument based on Barrett's Taxonomy. Key issues and concerns relating to the comprehension question taxonomies are raised.

**Chapter Four** uses ideas from Chapter Two and Three to discuss key features of a framework for assessing and comparing the cognitive challenge of Home Language examination so as to identify key criteria for instrument(s). Barrett's Taxonomy is compared with the Revised Bloom's Taxonomy. The usefulness of the two taxonomies to evaluate coverage of a variety of types of cognitive demands in Home Language examination papers is considered to see whether the two taxonomies might be combined to function more favourably. The Chapter discusses ways of addressing other factors that affect the level of difficulty of Home Language examination questions for Grade 12 students.

# Chapter 2: Question Taxonomies: Approaches, Debates and Concerns

## 2.1  Introduction

The main purpose of this chapter is to

- discuss educational taxonomies used for categorizing the cognitive challenge of questions; and

- identify the areas of concern or difficulty in taxonomies and instruments used to classify examination questions, including Umalusi's Home Language instrument based on the Revised Bloom's Taxonomy.


Morgan and Schrieber (1969 in Vogler, 2005: 98) define question taxonomies as hierarchical models used to classify questions in terms of increasing complexity 'based on the intellectual or mental activity needed to formulate an answer'. They are constructs used to classify questions based a) on the kind of knowledge students need to answer them; and/ or b) what students are required to do with knowledge in order to formulate an answer.

Although there are a number educational taxonomies based on hierarchical models of increasing complexity, not all are designed or appropriate for evaluating the cognitive challenge of examination questions. For example, the SOLO (Structure of the Observed Learning Outcomes) Taxonomy (Biggs and Collis, 1982) is designed in accordance with Piagets' different cognitive stages of development. However, its purpose is to categorize student **responses** in terms of pre-structural, uni-structural, multi-structural, relational and structural knowledge, as opposed to classifying the level of cognitive challenge of **questions**. The first level, pre-structural knowledge, for example, represents student responses that are irrelevant or that miss the point (Biggs, 1999).

In terms of educational taxonomies used for categorizing the cognitive challenge of questions, Bloom's Taxonomy is the best known.

## 2.2  Bloom's Taxonomy of the cognitive domain

Bloom's Taxonomy (Bloom et al., 1956) comprises three parts or domains of educational activity – the cognitive (knowledge and mental skills including processing information), affective (attitudes, feelings and values) and psychomotor (motor skills/manual or physical skills) domains (Krathwohl, Bloom & Masai, 1964).

For the purposes of this research report, the interest is in the cognitive domain which relates to intellectual activities and mental skills.

Essentially, Bloom's Cognitive Domain Taxonomy was developed as a hierarchical (as opposed to a relational) learning model for classifying educational objectives. The taxonomy comprises six categories of 'increasing levels of complexity and abstraction' (Hess et al., 2009:1). These categories, which are described as nouns, are ordered on a continuum from 'lower order' to 'higher order' processes in terms of increasingly complex or abstract mental activities.

The categories are: 'Knowledge', 'Comprehension', 'Application' (considered three lower levels of cognitive processes) and 'Analysis', 'Synthesis' and 'Evaluation' (considered three upper levels of cognitive processes with 'Evaluation' as the highest level).

Underpinning the model is the idea that teachers need to structure a series of learning experiences so that students can master each level before progressing to the next level. Each level has *instructional* implications. The 'lowest level' in the taxonomy – the 'knowledge' category involves students in understanding or memorising simple information such as facts, terms, and definitions. If only recall or recognition of basic knowledge is required, 'lecture' or reading may be adequate for learning, and teachers mainly need to ask students questions which require them to memorise, recall and reproduce simple information or material.

The idea is that, in order to progress beyond recall to 'Comprehension', students need to understand facts, information, or ideas, and interpret or infer meaning from material. To progress to the third 'Application' level, students need to be given opportunities to apply knowledge (known facts, rules, procedures, or methods, etc.) to problems. At the *instructional* level the implication is that they require some form of practice through the use of problems. If the learning process is initiated at higher taxonomic levels, then lower level processes need to be carefully scaffolded within tasks. For this reason the taxonomy has also become a tool in textbook development (Booker, 2007:352).

This taxonomy is also used to provide insight into the type of questions teachers ask in classroom instruction. It is used to provide a framework for guiding teachers in formulating questions, and in selecting and designing tasks that ensure that a variety of types of cognitive processes are included, especially in contexts where there are concerns about teaching which is predominantly characterized by reproduction of information, and replication of procedures. A framework such as Bloom's Taxonomy is thus considered useful as a guide for extending teaching repertoires.

Bloom's Taxonomy has been applied in various other ways in education including as a tool to plan examination questions so as to ensure that assessment includes a range of lower- and higher-level cognition questions or tasks. The taxonomy is used to evaluate appropriate coverage of a variety of types of cognitive demands in examination papers (Champeau De Lopez et al., 1997). For example, using the taxonomy, Bloom found that at the *college* level over 95% of test questions required students to think at the 'lowest possible level', that of the recall or recognition of information rather than upper levels of cognitive processes (Hess, 2005:1).

Examples of higher level processes include:
- 'Analysis' tasks or questions that require students to break down a passage or material into its component parts, or to recognize patterns and understand relationships between the parts;
- 'Synthesis' tasks which require students to bring together parts or elements to form a whole, or to engage in creating a new or unique product or answer; and
- 'Evaluation' tasks which involve students in making comparisons, judgments and choices, for example, about the value of material or text or about solutions to a problem (Hosseini, 1993).

The strength of Bloom's Taxonomy is that it emphasizes 'the complexity of the cognitive activities involved in learning' and suggests that different levels of cognitive demands must

be taught and evaluated (Champeau De Lopez et al., 1997:33). Bloom's Taxonomy and others like it thus provide a framework for extending examination and assessment repertoires beyond factual recall and routine procedures by ensuring that questions cover subject content at different levels of cognitive demand.

The attraction of Bloom's Cognitive Domain Taxonomy is its simplicity – it is easy to understand and remember. Outwardly it appears to be relatively easy to use (Hancock, 1994). Nevertheless, the taxonomy does have its critics and limitations.

## 2.2.1 Debates and concerns: 'Knowledge' as the lowest level

In relation to Bloom's hierarchy of cognitive levels, a key critique which relates to instruction, but which is also relevant in relation to the ordering or hierarchy of levels for a cognitive challenge framework, relates to the notion that the category 'Knowledge' (facts, figures, information) forms the 'lowest level'. Although Bloom intended to imply that knowledge acts as the foundation for higher order processes, and that students must first have acquired the pre-requisite knowledge, comprehension, and application skills in order to be able to engage in analysis, synthesis and evaluation types of questions (Paul, 1993), some researchers argue that positioning 'Knowledge' at the lowest level sends the wrong signal.

Wineburg and Schneider (2010:61), for example, argue that this positioning suggests that 'knowledge' is teaching and learning's lowest aim, whereas the pinnacle of 'intellectual activity' is 'new knowledge' or knowledge creation. They say that shifting the 'Knowledge' category to the highest level would signal that 'the world of ideas is not fully known'.

Other authors argue that the positioning of 'Knowledge' in the taxonomy serves to encourages the view that disciplinary knowledge is less important (than general knowledge, for instance), when in reality writing tasks in a particular subject involve disciplinary knowledge as well as different levels of cognitive demand (Booker, 2007).

Authors such as Booker (2007) argue that the taxonomy has been used to promote the idea that 'higher order thinking' can be taught and assessed 'in isolation of specific content', when, in fact, Bloom's brief was not content-free (Booker, 2007:352). Rather, Bloom considered content knowledge to be 'the basis for all higher thinking'.

Booker (2007:349) points out that Bloom's Taxonomy was originally developed as a tool for college educators at the tertiary level rather than at the general school level. He attributes the current 'misappropriation' of Bloom's Taxonomy at the general school level to the 'dominant constructivist paradigm', which argues for teaching 'higher order thinking skills' such as analysing and evaluating, before students have 'mastered the fundamentals' (Booker, 2007:354). He believes that the use of Bloom's Taxonomy (and others like it) at the general school level, has downgraded the acquisition of domain knowledge. It has shifted the focus from 'the transmission of knowledge' (Booker, 2007:353). As a result, school students are not developing the disciplinary knowledge they need to reach 'higher ground' in tertiary education (Booker, 2007:348). He states that whilst 'good education aims at more than merely recalling facts' (Booker, 2007: 353), critical reasoning requires solid knowledge and understanding of 'the facts of the discipline' (Booker, 2007:355). More challenging demands can only follow once the basics have been taught and learnt.

Of course, the counter argument to this view is that it is students in high poverty contexts who are less likely to be exposed to more cognitively challenging forms of questions and tasks. The argument is that it is students in schools serving communities with low socio-economic

status (SES,) who are typically exposed to instruction that does not progress beyond the level of memorization and recall, and where the focus is on questions that have answers explicitly stated in the text. Students are simply required to extract and reproduce their responses directly from the reading passage (Allington, 1991; Moll, 1991).

The discussion about the location of 'knowledge' in a hierarchical model of cognitive demand is relevant in the current South African policy context, where there has recently been a clear shift from an outcomes-based education (OBE) system towards a knowledge-based system. The overarching critical outcomes underpinning South Africa's previous outcomes-based school curriculum, Curriculum 2005 (C2005), emphasized the development of higher order process skills such as analysing, evaluating, and creativity (Taylor & Vinjevold, 1999).

Changes to the South African school curriculum now give much more emphasis to the development of subject specific or disciplinary knowledge. Documents provide greater specification of the content that students are expected to cover in each grade. Ironically this shift largely arose out of concern that students in schools serving historically disadvantaged communities were not being provided with access to the specialized disciplinary knowledge needed to succeed in tertiary education.

Indeed, because of its vagueness in the area of content, it could be argued that, by emphasizing high levels of cognitive demand without projecting content knowledge as a dimension of cognitive challenge, South Africa's OBE curriculum promoted the teaching of process skills in isolation of content. Of further concern was that the outcomes-based model of learning tended to discourage the idea of memorisation by describing it as 'rote learning'. Yet, remembering is a crucial process in the acquisition of knowledge.

We return to this discussion and its implications for the development of a cognitive challenge framework for evaluating examination papers later in the report. Section 2.3 discusses the taxonomy that was selected by Umalusi as the basis for developing an instrument for evaluating the levels cognitive demand of examination papers, the Revised Bloom's Taxonomy.

## 2.3  The Revised Bloom's Taxonomy

Since the 1950s there have been various attempts to build on Bloom's earlier work by refining and revising Bloom's Taxonomy, or developing new classification schemes for assessing cognitive demand (Booker, 2007). The Revised Bloom's Taxonomy (Anderson et al., 2001) uses the same number of categories as Bloom's Taxonomy. However, there are key differences between the two taxonomies.

First, the Revised Bloom's uses verbs rather than nouns for each of the cognitive process categories. Whereas Bloom used the term 'Knowledge' rather than 'Recall', in the Revised Bloom's Taxonomy, 'Knowledge' was replaced with the verb 'Remember', 'Comprehension' was renamed 'Understand', 'Application', 'Analysis' and 'Evaluation' were changed to 'Apply', 'Analyse', and 'Evaluate', and 'Synthesis' was renamed 'Create'.

Secondly, the order on the taxonomy continuum changed slightly. The authors considered 'Creativity' to be higher within the cognitive domain than 'Evaluation', so 'Synthesis' changed places with 'Evaluation', and 'Evaluation' was no longer the highest level. The 'rearranged' categories are:

**Lower order processes**
- Remember
- Understand
- Apply

**Higher order processes**
- Analyse
- Evaluate
- Create

Table 2 provides a comparison of cognitive process levels in the two taxonomies and the main verb cues that are usually associated with questions assigned to each cognitive process (Anderson et al., 2001).

**Table 2.1: Comparison of Descriptors: Bloom's Original Taxonomy and the Revised Bloom's Taxonomy Cognitive Process Dimensions**

| Bloom's Taxonomy | The Revised Bloom Process Dimensions |
| --- | --- |
| Knowledge<br>Define, duplicate, label, list, memorize, name, order, recognize, relate, recall, reproduce, state | Remember<br>Retrieve knowledge or relevant information from long-term memory or material, recognise, recall, locate, identify |
| Comprehension<br>Classify, describe, discuss, explain, express, identify, indicate, locate, recognize, report, restate, review, select, translate | Understand<br>Determining meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarise, generalize, infer a logical conclusion (such as from examples given), predict, compare/contrast, match like ideas, explain, construct models (e.g. cause-effect) |
| Application<br>Apply, choose, demonstrate, dramatize, employ, illustrate, interpret, practise, schedule, sketch, solve, use, write | Apply<br>Carry out or use a procedure in a given situation; carry out (apply to a familiar/routine task), or use (apply) to an unfamiliar/non-routine task |
| Analysis<br>Analyse, appraise, calculate, categorize, compare, criticise, discriminate, distinguish, examine, experiment, explain | Analyse<br>Break (material/ information) into constituent parts, determine how parts relate to one another and to an overall structure or purpose, differentiate between relevant-irrelevant, distinguish, focus, select, organise, outline, find coherence, deconstruct (e.g. for bias or point of view) |
| Synthesis<br>Rearrange, assemble, collect, compose, create, design, develop, formulate, manage, organise, plan, propose, set up, write | Evaluate<br>Make judgments based on criteria or standards, check, detect inconsistencies or fallacies, judge, critique |
| Evaluation<br>Appraise, argue, assess, choose, compare, defend, estimate, explain, judge, predict, rate, core, select, support, value, evaluate | Create<br>Put elements together to form a novel, coherent whole or making an original product, reorganise elements into new patterns/structures, generate, hypothesize, design, plan, construct, produce for a specific purpose |

Adapted from Hess et al., 2009:2

Table 2.1 shows that, unlike the original Bloom's which begins with 'Knowledge', the Revised Bloom's Taxonomy begins at the level of 'Remember'. However, the initial level in both taxonomies requires students to recall and produce information or facts from memory, or recognize and extract responses from material. Both taxonomies consist of five other levels, each level representing more complex and demanding cognitive processes, and subsuming those levels below it.

Crucially, unlike the original taxonomy, the Revised Bloom's acknowledges an interaction between 'type of knowledge' and cognitive processes. Whereas Bloom's Cognitive Domain Taxonomy consists of one dimension, the Revised Taxonomy introduces a second dimension. It tries to separate out a knowledge dimension from a cognitive process dimension.

This knowledge dimension consists of a hierarchy of four levels each representing different types of knowledge (nouns) – factual, conceptual, procedural, and metacognitive knowledge. The idea is that 'type of knowledge' forms an important aspect of cognitive complexity. With this in mind, the researchers (see for example, Hugo, et al., 2008) present these levels on a matrix (Table 2.2 below) whereby cognitive processes (six levels) and different types of knowledge (four types) can be interrelated.

**Table 2.2: Revised Bloom's Taxonomy showing a 'matrix' of knowledge and cognitive process dimensions**

| Knowledge dimension | Cognitive process dimension | | | | | |
|---|---|---|---|---|---|---|
| | Remembering (or Recognizing) | Understanding | Applying | Analysing | Evaluating | Creating |
| Factual knowledge | | | | | | |
| Conceptual knowledge | | | | | | |
| Procedural knowledge | | | | | | |
| Metacognitive knowledge | | | | | | |

- Factual knowledge refers to the basic elements of knowledge that students must know to be acquainted with a discipline and solve problems in the discipline.
- Conceptual knowledge refers to knowledge of the interrelationships among the basic elements within a larger structure that enable them to function together.
- Procedural knowledge refers to knowledge of how to do something, methods of enquiry, and criteria for using skills, techniques, and methods.
- Metacognitive knowledge refers to cognition (reasoning) in general and awareness of one's own reasoning (Krathwohl, 2002: 215).

Like Bloom's, the Revised Bloom's has its critics and limitations.

## 2.3.1 Debates and concerns: Distinguishing level of difficulty

Some researchers question the order of the levels of cognitive processes in the Bloom's and the Revised Bloom's Taxonomies. The Revised Taxonomy, like the original Bloom, implies that each more advanced level subsumes the levels below it. In other words, the implication is that students cannot understand a concept (conceptual knowledge), unless they remember it (factual knowledge); and they cannot apply a procedure (procedural knowledge) if they do not remember and understand it.

Researchers such as Newman (1993) and Paul, (1993) argue that students cannot necessarily recall knowledge without first understanding it or, that students do not necessarily need to understand a procedure in order to apply it.

Other authors (Guttman, 1953; Sax, Eilenberg, and Klockars, 1972; Seddon, 1978) argue that cognitive levels on their own do not necessarily distinguish between degrees of difficulty of questions. They point out that questions categorized at higher cognitive levels in the taxonomies are not necessarily more difficult than other questions categorized at the same level, or at lower levels.

For example, memorization questions can ask students to recall or recognize simple or much more complex information (factual knowledge), procedures (procedural knowledge) or concepts (conceptual knowledge). 'Application' or 'Understanding' questions can ask students to apply or explain simple or much more complex information, procedures or concepts. This argument implies that levels of difficulty are distinct attributes within each cognitive level.

The introduction of a separate knowledge dimension, naming four different types of knowledge that interact with cognitive processes in the Revised Bloom's, does not necessarily solve the problem of distinguishing 'levels of difficulty'. Researchers question the assumed hierarchical nature of both the cognitive process and 'type of knowledge' taxonomies.

Arguably, by their nature, remembering and understanding questions require factual and conceptual knowledge (as opposed to procedural and metacognitive knowledge), applying requires the use of factual, procedural and (ideally) conceptual knowledge; whilst analysing; evaluating and creating generally require the use of factual, conceptual, procedural and (possibly) metacognitive knowledge.

Finally, although the categories of the cognitive process taxonomies for both Bloom's and the Revised Bloom's (and the 'type of knowledge' taxonomy in the Revised Bloom's) are intended to transcend subject matter content (Hancock, 1994), a criticism is that the ordering or hierarchy of the cognitive process levels is not the same for different subjects.

Researchers such as Phillips and Kelly (1975) argue that cognitive complexity proceeds in a different order depending on the subject.

The instrument used by Umalusi in the analysis of the 2010 Home Language examination papers was developed specifically for assessing Language examination papers. The taxonomy used is more sophisticated than the Revised Bloom's Taxonomy in that it tries to distinguish between levels of difficulty within different cognitive processes.

The instrument aims to reflect the fact that each category of cognitive processes can

be tested at different levels, and that the cognitive levels on their own do not necessarily indicate level of difficulty.

## 2.4 Umalusi's Home Language instrument based on the Revised Bloom's Taxonomy

Table 2.3 below presents Umalusi's Home Language examination paper analysis instrument based on the Revised Blooms.

The two variables that Umalusi uses in order to determine the standard of Home Language examinations are: 'type of cognitive demand'; and 'degree of difficulty' of the cognitive process. There are five categories of cognitive demand. Within the different categories of cognitive demand on the grid presented below [namely, 'Conceptual knowledge', 'Comprehension (Understanding)'; 'Application'; 'Analysis and Problem-solving'; 'Evaluation and Synthesis (Creation)]', there are three different difficulty levels – 'Easy', 'Moderate', and 'Difficult'.

Teams of evaluators using the instrument to analyse the Home Language papers, first decide on the *type of cognitive* demand required to answer a particular examination question, and then decide on the *degree of difficulty* (as an attribute of the cognitive level) of the given examination question. In other words, the difficulty levels allow evaluators to distinguish between questions or tasks which are categorized at the same cognitive demand level, but which are not of comparable degrees of difficulty.[5]

**Table 2.3: Typology of cognitive demand used by Umalusi for Home Language examination analysis (based on Revised Bloom's Taxonomy)**

| Type of Cognitive Demand | Difficulty Level | Explanation and examples of level |
|---|---|---|
| **1. Conceptual knowledge**<br>- Recall and recite knowledge<br>- Remember, define and describe basic facts<br>- Identify, label, select, locate information<br>- Straight recall<br>- Identifying from text<br>- Know and use of appropriate vocabulary | **Easy** | Very simple recall; identify specific data; tell; recite; list e.g. identify parts of speech; match known words with definitions |
| | **Moderate** | Medium content, read and locate, briefly define a term, name and match. e.g. identify answers to wh- (equivalent) questions from a text; explain what synonyms are |
| | **Difficult** | Recall complex content; correct spelling and use of vocabulary. e.g. dictation of unfamiliar text; find synonyms or antonyms for words used in a text |

---

[5]The item-by-item analysis of the examination questions is used to determine which cognitive demand is most heavily weighted in papers, and whether papers lean towards a particular degree of difficulty.

| Type of Cognitive Demand | Difficulty Level | Explanation and examples of level |
|---|---|---|
| **2. Comprehension (Understanding)**<br>- Understanding of previously acquired information in a familiar context,<br>- Regarding information gathering: change or match information,<br>- Regarding use of knowledge: distinguish between aspects, compare and predict, defend and explain | **Easy** | Simple relationships; simple explanations; 1 step answers; e.g. converting active to passive forms; identifying main ideas (and supporting ones) in paragraphs; identify cause, result, reason from a text |
| | **Moderate** | Counter-intuitive relationships; give examples, explain, briefly summarise, translate, interpretation of realistic visuals. Summarise a text; draw inferences from a text or make a prediction |
| | **Difficult** | Identify principles which apply in a novel context; explaining; more complex reasoning with regard to understanding and explanation, motivate inferences or predictions made, using information from the text to support the position |
| **3. Application**<br>- Interpret and apply knowledge;<br>- Choose, collect and do basic classification of information;<br>- Modify existing text by making use of the existing knowledge<br>- Candidates to decide for instance on the most appropriate procedure to solve the solution to the question and may have to perform one or more preliminary calculations before determining a solution<br>- Select the most appropriate information from options<br>- Decide on the best way to represent data to create a particular impression | **Easy** | Write texts related to familiar contexts. Candidates know what process is required to solve the problem from the way the problem is posed. All of the information required is immediately available to the candidate; e.g. draft a friendly letter, basic business letter, invitation; given the necessary information, organize it into a presentable poster or a table to promote ready comprehension |
| | **Moderate** | Draw for instance information from given text; illustrate in words, construct ideas; e.g. propose a course of action based on a straightforward case study |
| | **Difficult** | Collect information from available texts to support a particular position/opinion and re-present the position in own text; e.g. undertake guided research to collect information necessary to a task; organise information into suitable form (report, memo, visual presentation) |
| **4. Analysis & Problem-solving**<br>- Analysis of information in a new or unfamiliar context / case study/ scenario;<br>- Examine and differentiate;<br>- Distinguish to find the most appropriate information;<br>- Research and investigate information Interpreting and extrapolating from solutions obtained by solving problems based in unfamiliar contexts | **Easy** | Simple process in known or practiced context, e.g. drafting an invitation, writing a letter of thanks or condolence, but with some variation which prevents the text from being strictly formulaic |
| | **Moderate** | Investigate, classify, categorise, compare, contact, solve, relate, distinguish, e.g. write a persuasive essay; take minutes of a straightforward meeting; deal with more complex case studies and propose course of action, e.g. in report form |

| Type of Cognitive Demand | Difficulty Level | Explanation and examples of level |
|---|---|---|
| **4. Analysis & Problem-solving**<br>- Using higher level cognitive skills and reasoning to solve problems<br>- Being able to break down a problem into its constituent parts – identifying what is required to be solved and then using appropriate methods in solving the problem<br>- Non-routine problems based on real contexts | **Difficult** | Complex abstract representation; referring to combination of concepts; Interpreting, report on, sort, debate, e.g. through preparing a speech and/ or presentation.<br>Using higher level cognitive skills and reasoning, e.g. in developing a proposal to solve a problem.<br>Being able to break down a problem into its constituent parts – identifying what is required to be solved and then using appropriate methods in solving the problem |
| **5. Evaluation & Synthesis (Creation)**<br>- Making judgment (evaluate), critique, and recommend by considering all material available;<br>- Weigh possibilities and do recommendations<br>- Construct new;<br>- Synthesise, create or find innovative solution;<br>- Formulate new ideas | **Easy** | Opinion, giving general critique on a fairly straightforward topic |
| | **Moderate** | Substantiate an opinion<br>Critique statements about situations made by others. Involving synthesis, critical argument; novel or abstract contexts; create poetry/a narrative |
| | **Difficult** | Generalise patterns observed in situations; working with complex problems involving insight and logic-leaps; creating new solutions to problems; redesign<br>Writing a complex review / critique<br>Re-write information / a story for a new context and setting |

Although this Umalusi instrument is based on the Revised Bloom's Taxonomy, it differs in fundamental ways.

First, the framework for Umalusi's instrument recognizes that a question that aligns to a particular type of cognitive demand is not always 'easier' than another question that aligns to the same cognitive demand level. For example, the instrument attempts to reflect the fact that a question can ask students to restate or explain a simple fact, or a much more abstract theory or complex content, the latter being much more difficult to accomplish. Low, medium and high order cognitive processes in the Umalusi framework each encompass the three levels of difficulty.

Secondly, where the Revised Bloom's has six levels of cognitive demand (Remember; Understand; Apply; Analyse; Evaluate; and Create), the Umalusi instrument has five. The taxonomy on the Umalusi instrument resembles the original Bloom's in that it uses nouns rather than verbs for each of the cognitive demand categories ('Application', 'Analysis' and 'Evaluation' rather than 'Apply', 'Analyse', and 'Evaluate'). The Revised Bloom's renamed 'Synthesis' 'Create', and made this the highest level within the cognitive domain rather than 'Evaluation', whereas the Umalusi instrument collapses these two categories into one level to become 'Evaluation and Synthesis (Creation)'. The Umalusi instrument has also added the term 'Problem-solving' to the 'Analysis' category.

The next section discusses some concerns about Umalusi's Home Language instrument.

## 2.4.1 Debates and concerns: Establishing degrees of difficulty

One concern about Umalusi's instrument is that, as Table 2.4 below illustrates, the notion of degree of difficulty does not necessarily correspond to the idea that a question that aligns to a particular cognitive process level is 'easier' than a question that aligns *to the level above it*. What Umalusi's framework implies, is that questions categorized at higher cognitive levels in the taxonomy are not necessarily more difficult than other questions categorized at lower levels.

**Table 2.4: Degrees of difficulty from Umalusi's Home Language instrument based on Revised Bloom's Taxonomy**

| Degree of difficulty | Explanation and examples of level |
|---|---|
| **EASY**<br>**1. Conceptual knowledge** | Very simple recall; identify specific data; tell; recite; list e.g. identify parts of speech; match known words with definitions |
| **2. Comprehension (Understanding)** | Simple relationships; simple explanations; 1 step answers; e.g. converting active to passive forms; identifying main ideas (and supporting ones) in paragraphs; identify cause, result, reason from a text |
| **3. Application** | Write texts related to familiar contexts. Candidates know what process is required to solve the problem from the way the problem is posed. All of the information required is immediately available to the candidate; e.g. draft a friendly letter, basic business letter, invitation; given the necessary information, organize it into a presentable poster or a table to promote ready comprehension |
| **4. Analysis & Problem-solving** | Simple process in known or practiced context, e.g. drafting an invitation, writing a letter of thanks or condolence, but with some variation which prevents the text from being strictly formulaic |
| **5. Evaluation & Synthesis (Creation)** | Opinion, giving general critique on a fairly straightforward topic |
| **MODERATE**<br>**1. Conceptual knowledge** | Medium content, read and locate, briefly define a term, name and match. e.g. identify answers to wh- (equivalent) questions from a text; explain what synonyms are |
| **2. Comprehension (Understanding)** | Counter-intuitive relationships; give examples, explain, briefly summarise, translate, interpretation of realistic visuals. Summarise a text; draw inferences from a text or make a prediction |

| Degree of difficulty | Explanation and examples of level |
|---|---|
| **MODERATE (continue)**<br>**3. Application** | Draw for instance information from given text; illustrate in words, construct ideas; e.g. propose a course of action based on a straightforward case study |
| **4. Analysis & Problem-solving** | Investigate, classify, categorize, compare, contact, solve, relate, distinguish, e.g. write a persuasive essay; take minutes of a straightforward meeting; deal with more complex case studies and propose course of action, e.g. in report form |
| **5. Evaluation & Synthesis (Creation)** | Substantiate an opinion. Critique statements about situations made by others. Involving synthesis, critical argument; novel or abstract contexts; create poetry/a narrative |
| **DIFFICULT**<br>**1. Conceptual knowledge** | Recall complex content; correct spelling and use of vocabulary.  e.g. dictation of unfamiliar text; find synonyms or antonyms for words used in a text |
| **2. Comprehension (Understanding)** | Identify principles which apply in a novel context; explaining; more complex reasoning with regard to understanding and explanation, motivate inferences or predictions made, using information from the text to support the position |
| **3. Application** | Collect information from available texts to support a particular position/opinion and re-present the position in own text; e.g. undertake guided research to collect information necessary to a task; organise information into suitable form (report, memo, visual presentation) |
| **4. Analysis & Problem-solving** | Complex abstract representation; referring to combination of concepts; Interpreting, report on, sort, debate, e.g. through preparing a speech and/or presentation. Using higher level cognitive skills and reasoning, e.g. in developing a proposal to solve a problem<br>Being able to break down a problem into its constituent parts – identifying what is required to be solved and then using appropriate methods in solving the problem |
| **5. Evaluation & Synthesis (Creation)** | Generalize patterns observed in situations; working with complex problems involving insight and logic-leaps; creating new solutions to problems; redesign. Writing a complex review / critique. Re-write information / a story for a new context and setting |

Unlike Umalusi's 'type of cognitive demand' taxonomy, the degrees of difficulty do not necessarily correspond with a hierarchical model of cognitive demand, and the notion that earlier levels are subsumed in the level above.

A 'difficult' 'Comprehension' question, for example, can be analysed as more difficult than an 'easy' 'Analysis and Problem Solving' question. *Identify principles which apply in a novel context; explaining; more complex reasoning with regard to understanding and explanation, explain inferences or predictions made, using information from the text to support the position* (i.e. a difficult Comprehension question) can be seen as more difficult than *Simple process in known or practiced context, e.g. drafting an invitation, writing a letter of thanks or condolence, but with some variation which prevents the text from being strictly formulaic* (i.e. An easy 'Analysis and Problem Solving' question).

Umalusi's instrument necessitates making judgments about each examination question at two points:

- first judging the type of cognitive demand; and

- then deciding on the degree of difficulty within that category of cognitive demand.

The finer categorization of 'Easy', 'Moderate' and 'Difficult' for each level of cognitive process makes it quite time-consuming to work with. The finer categorization also seems to require elaboration and (expanding) explanation and examples for each level making 'making judgments' complex.

Indeed a concern arose during the observation of the application of Umalusi's instrument by Umalusi's 2011 teams of Home Language examination evaluators in actual examination papers. This concern is the extent to which the various degrees of difficulty are interpreted uniformly by different evaluators, and how similarly questions are classified by evaluators in different language teams.

Evaluators appear to struggle to agree on the classification of questions to the different levels. This process tends to distract them from the idea that, what they are being asked to do, is to use their expertise and experience to make judgments about whether questions make low, medium or high demands on Grade 12 students.

The danger is that they may lose sight of the fact that one of the main purposes of determining whether an appropriate range of cognitive challenge is evident in the papers is to discriminate between high-achieving students, students performing at low levels, and students performing at levels in between.

The Umalusi instrument is based on the Revised Bloom's Taxonomy, but is specifically designed for assessing Home Language examination papers. As discussed in Chapter One, using different instruments and cognitive demand categories for different subjects makes it difficult to make comparisons among different subjects. There is a strong argument, therefore, that an approach to assessing 'cognitive challenge' needs to be redefined for each discipline.

Some researchers argue that differences between subjects lie in the categories of cognitive demand and their ordering. Other researchers argue that 'content complexity' forms an important dimension of cognitive demand, and that the difference in cognitive challenge lies mainly in the level of subject content knowledge needed to answer a question or perform a task.

For example, authors such as Stanovich (1986) and Todd, Tindal and Mccleery (2006) argue that it is a student's knowledge of a particular subject that also affects the difficulty level of a question. Anderson and Sosniak (1994) point out that a students' inability to evaluate, for example, may merely reflect the student's unfamiliarity with the particular topic, rather than a lack of evaluation skills.

These authors recognize that, unless students have necessary knowledge in a specific subject area, they will have a difficult time synthesising, analysing, or evaluating. In other words, they argue that students need particular levels of domain knowledge 'to think with'.

The **Revised Bloom's Taxonomy** separates out a knowledge hierarchy from a cognitive process hierarchy and categorizes both the cognitive processes and the type of knowledge (factual knowledge, conceptual knowledge, procedural knowledge, and metacognitive knowledge) required to answer a question or perform a task. However, the categories of cognitive processes and types of knowledge are 'generic'. The assumption is that categories in both dimensions apply 'equally' to different subject matter content and grade levels.

Instead of a separate knowledge dimension, **Umalusi's Home Language instrument** re-inserts 'Knowledge' as a cognitive process category. Whereas the Revised Bloom's Taxonomy of cognitive processes begins at the level of recall or recognition with the category 'Remember', Umalusi's instrument reverts to Bloom's original 'Knowledge' level, but uses one of the Revised Bloom's 'type of knowledge' categories, 'Conceptual Knowledge'.

A criticism of Bloom's, the Revised Bloom's Taxonomy (and by implication Umalusi's Home Language instrument based on the Revised Bloom's) is that these taxonomies do not relate cognitive challenge to the hierarchy that lies within the vertical knowledge structure of the subject area, especially in Mathematics, Science or Language (Bernstein, 1999). Some researchers (see for example, Muller, 2001) argue that there needs to be an entirely different language for assessing cognitive challenge and that this language should reflect the content complexity demands of examination questions within each particular subject area.

These researchers argue that content or topic complexity forms the key dimension of cognitive demand since engagement with more complex subject content is more demanding than engagement involving simpler subject content, and increases the complexity of a task or question (Stodelsky, 1988; Soden & Pithers, 2001). An example of a taxonomy that emphasizes the conceptual structure of the subject is the framework of Mathematics content outlined in Table 6 below (from Reeves & Muller, 2005).

The framework reflects different levels of 'content complexity' by showing grade level curriculum expectations for Intermediate Phase Mathematics in the National Curriculum Statements (DoE, 2002). Grade 4, Grade 5, and Grade 6 in brackets indicates elements that are considered as at the Grade 4, 5 or 6 levels. This particular extract describes grade levels for Number, operations and relationships: Recognizing, classifying and representing numbers: Representing and comparing whole numbers including zero and fractions.

**Table 2.5: Example of grade level 'content complexity' in Mathematics**

| | Representing and comparing whole numbers including zero and fractions including: |
|---|---|
| | Whole numbers to |
| 1 | 4-digit numbers (Grade 4) |
| 2 | 6-digit numbers (Grade 5) |
| 3 | 9-digit numbers (Grade 6) |
| | Common fractions with different denominators including |
| 4 | halves (Grade 4) |
| 5 | thirds (Grade 4) |
| 6 | quarters (Grade 4) |
| 7 | fifths (Grade 4) |
| 8 | sixths (Grade 4) |
| 9 | sevenths (Grade 4) |
| 10 | eighths (Grade 4) |
| 11 | tenths (Grade 6) |
| 12 | twelfths (Grades 5 & 6) |
| 13 | hundredths (Grade 6) |
| | Using operations appropriate to solving problems involving: |
| | Rounding off to the nearest |
| 48 | 10 (Grade 4) |
| 49 | 100 (Grade 4) |
| 50 | 1 000 (Grade 4) |

Table 2.5 provides a framework for assessing the 'level of content knowledge' required to answer a question or to complete a Mathematics task by indicating the grade level for which the content is appropriate. The framework allows evaluators to establish whether or not the level is consistent with (or lower than, or higher than) the content spelled out in the curriculum for the particular grade.

However, a content complexity 'taxonomy', such as the one above, listing all the various topics and sub-topics at different grade levels within a school phase for a subject area involves creating a very long and complex instrument. To make it easier for evaluators to assess the content level of a particular Mathematics task given to Grade 6 students, Reeves (2005) reduced the content complexity taxonomy to five levels. Evaluators used the following scale on a continuum from 1-5 to rate levels of content demands from (1) low to (5) high levels.

**Table 2.6: Levels of content complexity of Mathematics tasks given to Grade 6 srudents**

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **CONTENT LEVEL** | The content is mostly below the Intermediate Phase level (i.e. Grades 1, 2 or 3 levels). | The content is mostly at the Grade 4 level. | The content is mostly at the Grade 5 level. | The content is mostly at the Grade 6 level. | The content is mostly higher than the Phase level. |

Level 1 on Table 2.6 reflects very low content demands for Grade 6 students (i.e. consistent with curriculum content below Grade 4 level); Levels 2 and 3 reflect low content demands for Grade 6 student (i.e. consistent with curriculum content at the Grade 4 or 5 level); Level 4 reflects content demands consistent with curriculum content at the Grade 6 level; Level 5

reflects very high content demands (i.e. consistent with curriculum content above the Grade 6 level).

Other researchers recognize content complexity as integral to the notion of cognitive challenge, but also recognize an interaction between the structure of the knowledge domain (the 'level of content knowledge') and the cognitive processes required to answer questions or complete tasks. They argue that a clear distinction between 'levels of subject content knowledge' and cognitive process levels is needed. They argue that a two dimensional 'content by cognitive demand' matrix allows evaluators to consider the combination of grade level content demand and the cognitive demand of questions or tasks.

In Porter and Smithson's (2001) research, for example, items in the assessed or examined curriculum were content analysed. A topic-by-cognitive demand matrix was created to conduct a curriculum alignment analysis so that a curriculum alignment index could be created.

Table 2.7 below provides an example of a matrix of content and cognitive demand levels (from Reeves, 2005) which allows evaluators to classify questions based on the grade level of content knowledge students need to answer them; and what students are required to do with that knowledge in order to formulate an answer.

## Table 2.7: Levels of content by cognitive demand

| COGNITIVE DEMAND | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **CONTENT LEVEL** | The content is mostly below the Intermediate Phase level (i.e. Grades 1, 2 or 3 levels). | The content is mostly at the Grade 4 level. | The content is mostly at the Grade 5 level. | The content is mostly at the Grade 6 level. | The content is mostly higher than the Intermediate Phase level. |
| **COGNITIVE LEVELS**[6] | There are no or very low levels of cognitive demand. Students are engaged with very low levels of skills or tasks such as drawing or colouring in. | The levels of cognitive demand require *conceptual* knowledge of Mathematics that never 'evolves' into procedural knowledge. Students are engaged at conceptual levels only and do not progress to engagement with Mathematics procedures. | The levels of cognitive demand require *procedural* knowledge of Mathematics. Students are engaged with routine Mathematics procedures. | The levels of cognitive demand require *procedural* and some principled knowledge of Mathematics. Students are engaged with Mathematics procedures and to some extent with knowledge principles. | The levels of cognitive demand require *principled* and procedural Mathematics knowledge. Students are engaged with Mathematics procedures and to a larger extent with principles. |

---

[6]In this taxonomy a rating of 1 or 2 for *cognitive level* indicates that students are not really engaged with specialized mathematical knowledge. Ratings 3, 4 or 5 involve engagement with specialized mathematical knowledge. A rating of 5 involves the reasoning or principles behind the use of specialized mathematical knowledge (higher order processes).

What the above instrument recognizes is that questions or tasks may have low cognitive demand, but the demand of the content may be high.

However, an approach which relates cognitive challenge to 'type of cognitive demand' and the hierarchy that lies within grade level curriculum expectations probably does not have application value for the Home Language examination papers. Essentially, *the Subject Assessment and Examination Guidelines* prescribe the content, format and structure of Home Language examination papers (for example the number of sections, length and type of texts/ reading selections, time and mark allocation).

Umalusi's examination evaluators assess whether papers comply with these *Guidelines* and whether Learning Outcomes and Assessment Standards as set out in the *National Curriculum Statements* are covered. Nevertheless, it seems that there is a need to take into account the difficulty of questions specifically for Grade 12 level students, by judging whether questions are easy, moderately challenging, or difficult, for the average Grade 12 student to answer. This issue is discussed more fully in Chapter Four.

Finally, a key concern with Bloom's, the Revised Bloom's Taxonomy and the taxonomy used in Umalusi's Home Language instrument is that these taxonomies do not include cognitive processes specifically involved in comprehending written text. They are not specifically designed for assessing questions measuring comprehension of written text.

South Africa's Language curriculum documents (see for example, DoE 2008b) advocate a text-based approach where all aspects of language learning and assessment, including the formal aspects of language (knowledge of grammar and vocabulary), are studied and examined through the comprehension and interpretation of (written and graphic) text rather than in isolation from texts (including multimedia texts, such as advertisements, and posters).

# 2.5  Conclusions – question taxonomies

What emerges from the above review of question taxonomies and the debates and concerns raised is that:

a)  there needs to be a language for assessing cognitive challenge that reflects the cognitive demands of examination questions within the particular subject area;  and

b)  cognitive levels on their own do not necessarily distinguish between degrees of difficulty of questions. There is also a need to take into account changes in cognitive challenge expectations across the different grade levels, and the difficulty of questions specifically for Grade 12 level students.

The purpose of the reading comprehension taxonomies that are presented in **Chapter Three** is to assess questions that measure the skills involved in comprehending written (and graphic) text.

# Chapter 3: Reading Comprehension Taxonomies: Approaches, Debates and Concerns

## 3.1  Introduction

The main purpose of this chapter is to

- discuss educational taxonomies used for categorizing the cognitive challenge of reading comprehension questions.

- identify the areas of concern or difficulty commonly recognized in taxonomies used to classify reading comprehension questions.

There are a number of educational taxonomies that refer specifically to reading comprehension.

For example, Herber (1978) derived categories when attempting to relate Bloom's categories to three levels of reading comprehension. These are:

- Literal Comprehension, or questions requiring students to recall or recognize information explicitly presented in reading material;

- Interpretive Comprehension, or questions asking for a paraphrase, inference, or explanation; and

- Applied Comprehension, or questions requiring background knowledge to elaborate, predict, or solve problems based on implicit information in the text (Champeau De Lopez et al., 1997).

The specific purpose of Barrett's Taxonomy (used by the Department of Basic Education to set Home Language examinations) is to assess questions that measure reading comprehension.

## 3.2  Barrett's Taxonomy

Barrett's Taxonomy (Barrett, 1976) comprises five main 'comprehension' levels as opposed to the Revised Bloom's six cognitive process levels. These are (1) Literal Comprehension; (2) Reorganization; (3) Inferential Comprehension; (4) Evaluation; and (5) Appreciation.

The taxonomy is also more detailed than the Revised Bloom's Taxonomy in that each level contains between four and eight sub-categories. For example, for the category Literal Comprehension, Barrett mentions recognition or recall of sequence and of cause and effect relationships.

Table 3.1 is a summarized form of Barrett's Taxonomy which shows that some of the same sub-categories (for example, main ideas, comparisons; cause and effect relationships; character traits) are repeated at more than one level.

**Table 3.1: Summarized form of Barrett's Taxonomy**

| Type of cognitive demand | What questions require of the student/ examinee |
|---|---|
| **1. Literal Comprehension**<br>**Recognition or recall of**<br>- details<br>- main ideas<br>- a sequence<br>- comparison<br>- cause and effect relationships<br>- character traits | To locate or identify any kind of explicitly stated fact or detail (for example, names of characters or, places, likeness and differences, reasons for actions) in a reading selection/text/material |
| **2. Reorganization**<br>- classifying<br>- outlining<br>- summarising<br>- synthesising | To organise, sort into categories, paraphrase or consolidate explicitly stated information or ideas in a reading selection/text/material |
| **3. Inferential Comprehension**<br>- main ideas<br>- supporting details<br>- sequence<br>- comparisons<br>- cause and effect relationships<br>- character traits<br>- predicting outcomes<br>- interpreting figurative language | To use conjecture, personal intuition, experience, background knowledge, or clues in a reading selection/text/material as a basis of forming hypotheses and inferring details or ideas (for example, the significance of a theme, the motivation or nature of a character) which are *not explicitly stated* in the reading selection/text/material |
| **4. Evaluation – Judgment of**<br>- reality or fantasy<br>- fact or opinion<br>- adequacy or validity<br>- appropriateness<br>- worth, desirability and acceptability | To make evaluative judgement (for example, on qualities of accuracy, acceptability, desirability, worth or probability) by comparing information or ideas presented in a reading selection/text/material using external criteria provided (by other sources/ authorities) or internal criteria (students' own values, experiences, or background knowledge of the subject) |
| **5. Appreciation**<br>-Emotional response to content<br>-Identification with characters<br>-Reactions to author's language use<br>-Imagery | To show emotional and aesthetic/ literary sensitivity to the reading selection/text/ material and show a reaction to the worth of its psychological and artistic elements (including literary techniques, forms, styles, and structuring) |

Essentially, Barrett's Taxonomy distinguishes between questions that test 'Literal Comprehension', where answers to questions are 'textually explicit' and thus fairly obvious; questions that involve 'Reorganization', where the answers are in the text but not quite as obvious; and questions where students need to use a combination of a synthesis of the content of material plus their own internal script (their personal background knowledge of a particular subject or topic, their intuition, experience and imagination) to come up with an answer (Pearson and Johnson, 1978 in Champeau De Lopez et al., 1997).

Arguably, the taxonomy attempts to distinguish between questions which require students to 'read the lines' (Literal comprehension and Reorganization), 'read between the lines' (Inferential Comprehension), and 'read beyond the lines' (Evaluation and Appreciation) (Gray, 1960).

Section 3.3 provides the framework used by the Department of Basic Education based on Barrett's Taxonomy.

## 3.3 The Department of Basic Education's Home Language framework based on Barrett's Taxonomy

As discussed in Chapter One, the Department of Basic Education's (DoBE) examiners and moderators are required to set and moderate Home Language examination papers as reflecting the proportions of questions at the various levels of cognitive demand, based on Barrett's Taxonomy.

*The Examination Guidelines, Grade 12, Literature, Paper 2* for Home Language draws on Barrett's Taxonomy to provide guidelines on the type of questions that will be set as reflected in Table 3.2 below.

**Table 3.2: Home Language Examination Guidelines – Barrett's Taxonomy**

| Level | Description | Question types |
|---|---|---|
| 1 | Literal (information in the text) | e.g. Name the …; List the…; Identify the …; Describe the…; Relate the … |
| 2 | Reorganization (analysis, synthesis or organisation of information) | e.g Summarise the main ideas…; State the differences' similarities… |
| 3 | Inference (engagement with information in terms of personal experience) | e.g. Explain the main idea…; What is the writer's intention …; What, do you think, will be… |
| 4 | Evaluation (judgments concerning the value of worth) | e.g. Do you think that…; Discuss critically … |
| 5 | Appreciation (assess the impact of the text) | e.g. Discuss your response …; Comment on the writer's use of language … |

Source: DoE, 2009a:23

*The Examination Guidelines, Grade 12, Literature, Paper 2* for Home Language specifies that 'using Barrett's Taxonomy, various types of questions will be set in such a way that **all** the cognitive levels are catered for in the following proportions: Levels 1 and 2: 40% of total marks; Level 3: 40% of total marks; Levels 4 and 5: 20% of total marks'.

What can be inferred from the structure of the proportions, is that 'Literal Comprehension' and 'Reorganization' are considered to reflect lower order cognitive processes. 'Inferential Comprehension' is considered to reflect middle order cognitive processes. 'Evaluation' and 'Appreciation' are considered to reflect higher order cognitive processes.

Although the types of cognitive demands and the hierarchical order of cognitive levels on Table 3.2 match Barrett's categories, the detailed sub-categories that the Barrett's Taxonomy provides are not included in the above framework. For example, for the category 'Reorganization', the sub-categories *classifying, outlining*, and *synthesising* are not indicated.

However, Appendix 1 of the *Examination Guidelines Grade 12 Literature: Paper 2, Home Language and First Additional Language, and Paper 1: Section D: Second Additional Language* (DoE, 2009a, pages 24-26) provides the following elaboration of '**types of questions**':

**1. Contextual Questions** (Language and Literature Papers):

Contextual Questions are set on a variety of selected texts (in the Language Paper) and on extracts from the prescribed texts (in the Literature Paper) to assess language competency and to gauge the extent to which Assessment Standards prescribed in the NCS have been achieved. The level of complexity depends on the level at which the Language is being assessed (i.e. HL, FAL or SAL).

**1.1 Literal**:
Questions that deal with information explicitly stated in the text.
- Name the things/people/places/elements …
- State the facts/reasons/ points/ideas …
- Identify the reasons/persons/causes …
- List the points/facts/ names/reasons …
- Describe the place/person/character ...
- Relate the incident/episode/experience …
- Etc.

**1.2 Reorganization**:
Questions that require analysis, synthesis or organisation of information explicitly stated in the text.
- Summarise the main points/ideas/ pros/cons/ …
- Group the common elements/factors …
- State the similarities/differences …
- Give an outline of …
- Etc.

**1.3 Inference**:
Questions that require a candidate's engagement with information explicitly stated in the text in terms of his/her personal experience.
- Explain the main idea …
- Compare the ideas/attitudes/ actions …
- What is the writer's (or character's) intention/attitude/motivation/reason …
- Explain the cause/effect of …
- What does an action/comment/attitude (etc) reveal about the narrator/writer/ character …
- How does the metaphor/simile/image affect your understanding …
- What, do you think, will be the outcome/effect (etc) of an action/situation …
- True/False questions
- Multiple Choice questions
- Choose the correct option (from a given list)
- Fill in the blanks (using contextual clues)
- Questions on visual and graphic literacy
- Etc.

**1.4 Evaluation**:

These questions deal with judgements concerning value and worth. These include judgements regarding reality, credibility, facts and opinions, validity, logic and reasoning, and issues such as the desirability and acceptability of decisions and actions in terms of moral values.

- Do you think that what transpires is credible/realistic/ possible …?
- Is the writer's argument valid/logical/conclusive …
- Discuss/Comment critically on the action/ intention/motive/attitude/ suggestion/ implication …
- Do you agree with the view/statement/observation/ interpretation that …
- In your view, is the writer/narrator/character justified in suggesting/ advocating that … (Substantiate your response/Give reasons for your answer.)
- Is the character's attitude/behaviour/action justifiable or acceptable to you? Give a reason for your answer.
- What does a character's actions/attitude(s)/motives show about him/her in the context of universal values?
- Discuss critically/Comment on the value judgements made in the text.
- Etc.

**1.5 Appreciation**:

These questions are intended to assess the psychological and aesthetic impact of the text on the candidate. They focus on emotional responses to the content, identification with characters or incidents, and reactions to the writer's use of language (such as word choice and imagery).

- Discuss your response to the text/incident/situation/ conflict/dilemma …
- Do you empathise with the character? What action/decision would you have taken if you had been in the same situation?
- Discuss/Comment on the writer's use of language …
- Discuss the effectiveness of the writer's style/ introduction/ conclusion/imagery/ metaphors/use of poetic techniques/ literary devices …
- Etc.

Appendix 1 of the *Examination Guidelines Grade 12 Literature: Paper 2 includes a second section headed '***the Literary Essay***', which states that:

- An Essay question requires a sustained piece of writing of a specified length on a given topic, statement, point of view or theme.
- The literary essay requires a candidate to discuss/discuss critically a comment/ statement/viewpoint on a specific text. The essay may be argumentative or discursive, and involves a candidate's personal response to and engagement with the text.

Before I address areas of concern or difficulty with comprehension taxonomies, Section 3.4 presents two other examples of taxonomies based largely on Barrett's Taxonomy.

# 3.4  Other Taxonomies based largely on Barrett's Taxonomy

An example of a 'simpler' taxonomy based on Barrett's is the Progress in International Reading Literacy Study (PIRLS) Taxonomy.

## 3.4.1 The Progress in International Reading Literacy Study (PIRLS) Taxonomy

The main dimension of the written test in the PIRLS assessment of the reading literacy of young children is reading comprehension.

Table 3.3 shows the PIRLS typology of comprehension processes and the tasks that exemplify each process.

**Table 3.3: PIRLS typology of comprehension processes and the tasks that exemplify each process**

| Type of comprehension process | Tasks that exemplify this process |
|---|---|
| Focus on and retrieve explicitly stated information | Identifying information that is relevant to the specific goal of reading<br>Looking for specific ideas<br>Searching for definitions of words and phrases<br>Identifying the setting of a story (e.g. time, place)<br>Finding the topic sentence or main idea (when explicitly stated) |
| Make straightforward inferences | Inferring that one event caused another event<br>Concluding what is the main point made by a series of arguments<br>Determining the referent of a pronoun<br>Identifying generalizations made in a text<br>Describing the relationship between two characters |
| Interpret and integrate ideas and information | Discerning the overall message or theme of a text<br>Considering an alternative to actions of characters<br>Comparing and contrasting text information<br>Inferring a story's mood or tone<br>Interpreting a real-world application of text information |
| Examine and evaluate content, language and textual elements | Evaluating the likelihood that the events described could really happen<br>Describing how the author devised a surprise ending<br>Judging the completeness or clarity of information in the text<br>Determining an author's perspective on the central topic |

Adapted from Mullis et al., 2009, Chapter 2

The PIRLS taxonomy for describing comprehension questions consists of four types of reading comprehension processes as opposed to Barrett's five levels.

The PIRLS Taxonomy classifies questions which require young children to

1. focus on and retrieve explicitly stated information (require reader to 'recognize the relevance of the information or idea in relation to the information sought') (Mullis et al., 2009:25)

2. make straightforward inferences (require the reader 'to fill in the "gaps" in meaning that often occur in texts') (Mullis et al., 2009:25)

3. interpret and integrate ideas and information (require the reader to 'process text beyond the phrase or sentence level' and 'be open to some interpretation based on their own perspective') (Mullis et al., 2009:27) and

4. examine and evaluate content, language and textual elements (require the reader to 'stand apart from the text and examine or evaluate it' by relying 'on knowledge about the world or past reading') (Mullis et al., 2009:28).

Barretts' five levels include: 1) Literal Comprehension; (2) Reorganization; (3) Inferential Comprehension; (4) Evaluation; and (5) Appreciation.

Because the PIRLS Taxonomy is orientated towards assessing the reading comprehension of primary school children, it does not cover the highest level of Barrett's Taxonomy – questions which require students to show emotional and aesthetic or literary appreciation. Instead of the category 'Reorganization', the PIRLS Taxonomy draws a distinction between questions requiring 'straightforward inferences', and questions which require more sophisticated levels of interpretation where the reader has to '*integrate ideas and information*'.

An example of a more complex instrument which draws on Barrett's Taxonomy is Umalusi's Home Language instrument used by the 2011 teams of Home Language examination evaluators.

## *3.4.2 Umalusi's Home Language instrument based on Barrett's Taxonomy*

Because of the differences between the categories used in Umulusi's instrument based on the Revised Bloom's Taxonomy and the categories used by the Department of Basic Education for setting the Language and Literature (Home Language) examination papers, Umalusi's 2011 teams of examination evaluators used the following instrument based on the Barrett's Taxonomy.

This instrument, which was developed for the comparative analysis of Home Language papers in 2011, includes the finer categorizations of 'Easy', 'Moderate' and 'Difficult' discussed in Chapter Two. The examples and explanations for each level had to be adjusted and adapted to align with Barrett's reading comprehension taxonomy levels.

**Table 3.4: Umalusi's Home Language instrument for Home Language examination analysis based on Barrett's Taxonomy**

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| **Literal Comprehension (LC)**<br><br>**To identify information directly stated**<br><br>Recognition and recall of:<br><br>• Ideas and information explicitly stated<br>• Details, main ideas<br>• Provide from memory explicitly stated reasons for actions | **Easy** | Identify main ideas. Simple recall answers; identify specific data; tell; recite; list e.g. identify parts of speech; Read and locate, briefly define a term, name – e.g. identify answers to wh- (equivalent) questions from a text | List the names of the characters involved… Skim read to identify… Tell the story… Define a term… What, where, when, who questions… What is the name of, for instance, a person, thing, place etc.; What happened at a particular place… What happened to shorten his stay… |
| • Sequence<br>• Recognition of comparison<br>• Cause and effect relationships<br>• Character traits | **Moderate** | Recall more complex content as a series of facts; simple relationships; simple explanations; identifying main ideas (and supporting ones) in paragraphs; identify cause, result, reason directly from a text | To find what a character did… Who, for instance, lived on the farm? When did something happen? Which words in the intro…? or Mention the word… How much land was claimed … How did they find …; How did they accomplish… Over what kind of land did they travel… (recall with no sequencing or reorganization) |
| | **Difficult** | Give examples, explain, briefly summarise, translate, interpretation of realistic visuals. Look at an illustration and tell the story in sequence; Summarise a text; identify principles which apply in a novel context; explaining; more complex reasoning with regard to understanding and explanation. Explain single cause and effect principles (What was the reaction of … to …). Identify and explain particular character traits | Identify the order of incidents… (general sequencing) Identify explicit statements… Identify likenesses / differences in characters… Look for ideas which conflict each other… Fill in the missing word(s)? Refer to a particular style indicator… (literature) Explain what happened with the main character. Explain the place or the main character… Indication of sequence… Do a flow diagram to indicate sequence… Complete the table to indicate the similarities and the difficulties Put words in order… Find a sentence that tells why… Find ways to explain the character… Finding a pattern; what happened first, second or last… Summarise her attitude toward life. (still recall of explicit statement) |

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| **Reorganization (R)** <br> **To organize or order the information in a different way than it was presented** <br><br> • Classifying <br> • Outlining <br> • Summarising <br> • Construct ideas <br> • Utilise ideas from the text <br> • Paraphrase or translate the author's statements | **Easy** | Write texts related to familiar contexts. <br> Candidates know what process is required to solve the problem from the way the problem is posed. | Classify ideas… Organize information… reorganize some facts… <br> Classify the following according to… <br> Placing people, things, events in categories…. <br> Place the following under the proper heading…. <br> Give a summary of…. / outline main ideas <br> Order ideas / information under a particular heading… <br> Divide the story according to particular parts…. <br> To relate ideas to a theme…. <br> Tell the story in your own words… Describe the tone, using your own words … (construct ideas) <br> Multiple-choice questions…. / Which of the following doesn't belong… <br> What are the similarities / differences between two or more characters, stories or poems…. <br> Group the common characteristics / factors / elements in table form or Venn-diagram… <br> Combine the information from different sources in a paragraph…. <br> Do a mind map to illustrate understanding; view; perspective… |
| | **Moderate** | Candidate to organize information into a presentable poster or a table to promote ready comprehension. | |
| | **Difficult** | Draw for instance information from given text; illustrate in words, construct ideas; e.g. propose a course of action based on a straightforward case study. Consolidate ideas from more than one source; discuss poetic devices such as repetition, symbolism. | |

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| **Inferential Comprehension (IC)** **To respond to information in a different way than it is presented** • Answers not explicitly stated – must be inferred • Inferring supporting details • Using intuition/ personal experiences • Thinking and evaluation that go beyond the printed page | Easy | Collect information from available texts to support a particular position/opinion and re-present the position in own text; e.g. undertake guided research to collect information necessary to a task; organize information into suitable form (report, memo, visual presentation) Simple process in known or practiced context, e.g. drafting an invitation, writing a letter of thanks or condolence, but with some variation which prevents the text from being strictly formulaic | Explain what is meant by… (answer not found in text) Explain / illustrate in own words… Answer a riddle… Write a sentence that summarise the main idea… Was the discovery planned or accidental…? Construct ideas based on what you've read … Provide reasons for your understanding of…. Predict what would happen / will be the result… Compare the characters / stories based on interpretation of features / character traits… Make inferences from the character's reaction / response … What is the main idea in the paragraph… the main theme in the story / poem… Explain the metaphor… What would possibly be the result / effect of …. What do you think will happen hereafter… Identify and explain what is implied… |
| • Organize main ideas in a suitable form • Counter-intuitive relationships, collect information from available texts to support a particular position/opinion and re-present the position • Interpreting figurative speech | Moderate | Investigate in more detail, establish what the present is revealing of the future, solve by reading between the lines, relate, distinguish between, e.g. write a persuasive essay; take minutes of a straightforward meeting; deal with case studies and propose course of action, e.g. in report form. explain what a character feels when expressing particular feelings. Discuss impact of e.g. rhetorical questions. | Identify the general significance, theme or moral not explicitly stated… React on implied information… Point to the suggested / possibly preferred reaction of a character…. Suggest consequences for the reaction of a character… What connotation can be made… and to substantiate such connotation from the text / source material… Multiple-choice questions where educated guesses are required …. Indicate whether the story from another era and the circumstances in the story / novel may occur in present day terms …. Journal writing for a particular character… Work with information which is suggested, insinuated, and to make inferences from the given text / source material… What would be the implications of… |

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| | Difficult | Complex abstract representation; referring to combination of concepts; Interpreting, report on, sort, debate, e.g. through preparing a speech and/or presentation. Using higher level cognitive skills and reasoning, e.g. in developing a proposal to solve a problem. Being able to break down a problem into its constituent parts – identifying what is required to be solved and then using appropriate methods in solving the problem. Find phrases to convey messages / impressions / implications. Discuss repercussion; provide substantiation. | To identify what the repercussions would be… Read and interpret an allegation…. Interpret what is hinted at… What ideas are brought to mind…/ What makes ____ a ____ What caused the author to include particular words, ideas, characterisations… What did ____ prove about his / her attitude toward____ What was ____'s attitude about____ Interpret the following figurative expressions… |

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| **Evaluation (E)**<br>**To make judgments in light of the material**<br><br>• Sorting fact from opinion, same/ different and good or bad. E.g. To read a story and decide which character is the most like their own personality. Evaluation and awareness questions to be answered: Why do you think so? How did you know?<br>• Judgment and focuses on qualities of accuracy and probability; appropriateness<br>• Comments based on judgments of moral character | Easy | Opinion; giving general critique on a fairly straightforward topic; general comments on style; evaluate effectiveness of image | Give an opinion whether what happened with the character may happen with a person in real life... Indicate whether... Is a fact or opinion. Provide reasons for the answer... |
| | Moderate | Evaluate in more detail; compare and substantiate choice; evaluate the use of poetic devices; evaluate effectiveness of image | Do you think the character's dialogue is realistic? Is the speaker's argument logic and / or valid? Critically evaluate the attitude / action of the character... Do you agree / disagree with the view/ perspective / interpretation... Do you think it's good for a character to pretend... give reasons for your answer... |
| | Difficult | Weigh possibilities and provide reasons; make recommendations; to provide adequate support for conclusions; comments on appropriate or effective use of e.g. metaphors; evaluative explanation of e.g. contradictions; comments on the accuracy of statements | What does the attitude / reaction of... suggest about the character's view of other people... Discuss a poem as example of eg a sonnet Defend why a particular short story can serve as an example of a successful short story... Propose ideas / make suggestions based on an evaluation... Could this really happen...? Is this fact or fiction? Give reasons for your answer... What strange ideas _____ have? Which ideas are still accepted and which are no longer believed? Evaluate the appropriateness of text... / appropriateness of figurative speech Is a character acting fairly... on what grounds would one make the claim |

| Barrett's taxonomy (based on interpretation and appreciation of text) | Degree of Difficulty | Explanation and examples per level of difficulty | Example questions |
|---|---|---|---|
| **Appreciation (A)**<br><br>**To give an emotional or image-based response** | **Easy** | General emotional response with little substantiation; identification with characters or incidents; general reactions. | Do you like the main character… Substantiate your view…<br>What part of the story did you find most exciting – provide reasons…<br>Do you think that he will follow the advice, keeping his personality in mind…<br>What would you do if you were _____ ?<br>Illustrate response in words… |
| • Emotional response to the content<br>• Identification with characters or incidents<br>• Reactions to the authors' use of language<br>• Imagery<br>• Response based on own opinion | **Moderate** | Substantiate an opinion Critique statements about situations made by others. Involving synthesis, critical argument; novel or abstract contexts; create poetry/a narrative. | Give commentary on a character's values and ethical choices…<br>Justify your answer<br>Respond on e.g. a dilemma or conflict in a story / poem…<br>On what grounds can you identify with the character…<br>Be able to defend the actions of a character…<br>Write a conclusion for … |
| • Critical review based on interpretation and appreciation | **Difficult** | Critical review based on the ability to interpret, illustrate in words and provide an appreciative opinion. Generalise patterns observed in situations; working with complex problems involving insight and logic-leaps; creating new solutions to problems; redesign. Writing a complex review / critique Re-write information / a story for a new context and setting. | Commentary on the appropriateness of figure of speech; language use…<br>Discuss the appropriateness of an introduction, close, style of writing etc.<br>Rewrite a part in a novel / a drama as eg a poem…<br>Give a substantiated appreciated opinion…<br>Propose new solutions…<br>Rewrite information / use information in a new applied setting…<br>Writing appreciative comments based on observation…<br>Critical reviewing processes…<br>Write a poem / short story on a particular topic… |

As discussed in Chapter Two, one concern with this framework is that the notion of degree of difficulty does not necessarily correspond with the notion that a question that aligns to a particular reading comprehension process level is 'easier' than a question that aligns to the level above it.

The finer categorization of degrees of difficulty on Umalusi's instrument makes it difficult to establish whether different evaluators are interpreting and classifying the three levels for each comprehension process category comparably or similarly.

The fourth column on this instrument is intended to provide examples of questions that link the comprehension process categories to examples, but linking examples to each of the three levels of difficulty for each process level in the taxonomy is complicated.

A key constraint with using Barrett's Taxonomy categories is that they are designed only to cover reading comprehension processes. Using the reading comprehension process categories on Barrett's Taxonomy to assess Home Language examination papers means that categorizing tasks such as 'drafting an invitation, writing a letter of thanks or condolence' becomes somewhat contrived.

Section 3.5 discusses this and other limitations in more detail.

## 3.5 Debates and Concerns: Difficulty level of texts and 'covering' creative writing

Conventionally Language examination papers (including South Africa's Grade 12 examinations) cover more than comprehension of text, and include aspects such as grammar and creative writing (with topics that do not necessarily include text-based stimuli).

A limitation of most reading comprehension taxonomies, including the Barrett's Taxonomy, in terms of evaluating the cognitive challenge of Language examination papers, is that they are not designed to evaluate questions that test grammatical skills in isolation of text (for example, grammatical rules), or tasks which require students to generate or create their own text (for example, a full composition or a letter), or other original product (such as designing an advertisement).

As Table 3.5 below shows, the Grade 12 Home Language examinations comprise four papers and Continuous Assessment (CASS).

**Table 3.5: Structure of Home Language examinations**

| Exam paper | Covered in the paper | Marks | |
|---|---|---|---|
| **Paper 1** | **Language in context** | | |
| | **Comprehension**: Passages(s), visuals/graphics, explanations, descriptions, etc. | 30 | **70** |
| | **Summary**: Summarising in point or paragraph form | 10 | |
| | **Language**: Advertisements, cartoons, prose texts, structures, critical awareness, purpose, reports, descriptions, procedures, explanations etc. | 30 | |
| **Paper 2** | **Literature** | | |
| | **Seen and unseen poetry**: Essay & contextual questions or 2 contextual questions + essay or contextual question (unseen poetry) | 30 | **80** |
| | **Novel**: Essay or contextual question | 25 | |
| | **Drama**: Essay or contextual question | 25 | |
| **Paper 3** | **Creative writing** | | |
| | **Essay**: Narrative, descriptive, reflective, argumentative, expository, or discursive | 50 | **100** |
| | **Longer transactional text**: Newspaper articles, diary, memoranda, minutes and agenda, letters, etc. | 30 | |
| | **Shorter text**: Flyers, instructions, advertisements, post cards, posters, etc. | 20 | |
| **Paper 4** | **Oral**: Reading, prepared or unprepared speech, conversation or debate, interview, etc. | | **50** |
| **CASS** | 14 tasks (900 divided by 9) | | **100** |
| **Total for Home Language Examination** | | | **400** |

As indicated, the examination comprises external and internal components, which together give a total of 400 marks. The external component (Papers 1 - 4) is marked out of 300 marks, of which 50 marks is constituted by the oral tasks undertaken during the year. The four papers contribute 75% to the final NSC mark. The remaining 25% is made up of Continuous Assessment (CASS), consisting of 14 tasks which are all internally assessed by schools.

Clearly a limitation of the categories on Barrett's Taxonomy is that:

(1) They do not necessarily address or cover all dimensions of Home Language examination papers. For example, the taxonomy does not include categories of cognitive demand that can easily be related to Paper 3: Creative writing (above). Barrett's Taxonomy is not designed to cover writing demands or tasks that require students to generate or produce their own text or material, for example, by using their imagination, etc.

(2) Not all Language examination papers are text-based, or include questions based on literary texts (for example Business English offered at vocational colleges). Most reading comprehension taxonomies describe or classify the cognitive demand of comprehension questions but make no reference to the difficulty level of the texts or material that students are supposed to read and the reading demands they make on students.

(3) Some of Barrett's sub-categories are more likely to be associated with questions based on different types of texts. For example 'interpreting figurative language', judgment of 'reality or fantasy', or recognition of 'character traits' are more likely to be associated with questions based on fictional or literary texts (for example, poetry, plays) than on informational or non-fictional texts (for example, newspaper articles, advertisements, graphs or tables) (Mullis et al., 2009).

However, although questions in an examination paper may be similar to questions in another paper in terms of the type of cognitive demand, the prescribed texts or reading passages (such as novels, plays and poems), and visual **texts selected may differ significantly over different years and across different Home Language examinations, in terms of their intrinsic difficulty.**

**This difference makes the examination more (or less) cognitively challenging. For example, literal comprehension of a simple contemporary magazine article is easier that literal comprehension of a classical work such as one of Shakespeare's plays.**

Thus authors such as Champeau De Lopez et al., (1997) argue that, in addition to categorizing the degree of challenge of the questions that are asked, it is also necessary to take into account the complexity or difficulty level of reading selections, texts, source or stimulus material (including prescribed novels, poems, and drama, the visual texts, for example, cartoons) selected for comprehension summary and language work in Language examination papers.

As Mullis et al. (2009:24) observe 'all texts are not equal', and texts can be of very differing standards or degrees of difficulty in terms of content, register, diction, abstractness of ideas, and imagery.

**Differences in the degree of challenge in prescribed set works across Home Languages, for example, can make an immense difference to the cognitive challenge of the different Home Language Literature examinations.**

# 3.6 Conclusion – reading comprehension Taxonomies

What emerges from this review of reading comprehension taxonomies, and the debates and concerns raised, is that:

a) any taxonomy selected or constructed for evaluating Home Language examination papers needs to cover all aspects of the examination papers, for example, literature, grammar, creative writing, etc.;

b) there is also a need to take into account changes in comprehension expectations at different grade levels, and the difficulty of comprehension questions specifically for Grade 12 level students;

c) the degree of difficulty of the text that has to be read has a material influence on the level of cognitive challenge – regardless of the type of questions asked about the text.

Chapter Four uses the approaches, debates or concerns that are raised in Chapters Two and Three to consider ways in which Umalusi's 'cognitive challenge' framework could be made more suitable for assessing Home Language examinations.

# Chapter 4: Establishing a 'Cognitive Challenge' framework for Home Language Examinations

## 4.1  Introduction

The main purpose of Chapter Four is to consider the approaches, debates and concerns outlined in Chapters Two and Three to:

- select the taxonomy most suitable for assessing cognitive demands in Home Language examination papers.

- determine the most appropriate ways of assessing the degree of difficulty of Home Language examination questions

## 4.2  Defining 'Cognitive Challenge'

It seems that there is no simple definition of what constitutes 'cognitive challenge'. What is clear, is that, for a range of cognitive challenges to be evident in Home Language examination papers, there needs to be

a)  appropriate coverage of a variety of types of cognitive demands (lower-, medium- and higher-level cognitive demands); and

b)  questions set at difficulty levels that make it possible to discriminate among students performing at high levels for the grade, students performing at low levels, and students performing at levels in between.

Thus, the relative proportions of examination questions at particular cognitive and difficulty levels are very important. However, the degree of difficulty of the text that has to be read also has a material influence on the degree of cognitive challenge.

What is also evident, is that

a)  a hierarchy of cognitive demands is a useful tool for planning examination questions so as to ensure that *assessment* includes a range of lower-, medium- and higher-level cognition questions or tasks;

b)  having an instrument with standardized criteria to *evaluate* appropriate coverage of a variety of types of cognitive demands in examination papers is useful for making judgments about the 'cognitive challenge' of examination papers as fair and rigorous as possible; and

c)  a taxonomy selected or constructed for evaluating Home Language examination papers needs to cover all aspects of the examination, for example, literature, grammar, creative writing, etc.

Clearly a detailed relatively low-inference rating for 'type of cognitive demand' is useful for checking that an appropriate range of cognitive demands is expected of students, and that higher level cognitive demands are not under-represented in examination papers.

The key issue is: **What taxonomy is the most suitable for assessing coverage of a variety of types of cognitive demands in Home Language examination papers?**

## 4.3 Selecting the Taxonomy most suitable for assessing Home Language examination papers

By now most Home Language teachers and Umalusi's expert evaluators are likely to be familiar with the Revised Bloom's Taxonomy and Barrett's Taxonomy. Both taxonomies have strengths that can complement the other's weaknesses.

However, both Barrett's and the Revised Bloom's Taxonomies vary in their particular strengths and weaknesses. The next section compares weaknesses and strengths of Barrett's Taxonomy, the Revised Bloom's Taxonomy and the taxonomy used in Umalusi's instrument, based on the Revised Bloom's Taxonomy.

### 4.3.1 Comparing Barrett's Taxonomy with the Revised Bloom's Taxonomy and the taxonomy used in Umalusi's instrument based on the Revised Bloom's Taxonomy

Table 4.1 provides the various categories of cognitive processes on the Barrett's and the Revised Bloom's Taxonomies and modified version of the Revised Bloom's Taxonomy used for Umalusi's Home Language examination instrument.

**Table 4.1: Categories of cognitive processes on the Barrett's and the Revised Bloom's Taxonomies, and the taxonomy on Umalusi's instrument based on the Revised Bloom's**

| Barrett's Taxonomy | Revised Bloom's Taxonomy | Type of Cognitive Demand according to Umalusi's instrument based on the Revised Bloom's |
|---|---|---|
| 1.Literal Comprehension | 1.Remember<br>2.Understand | 1.Conceptual knowledge |
|  |  | 2.Comprehension (Understanding) |
| 2.Reorganization | 3.Apply<br>4.Analyse | 3.Application |
|  |  | 4.Analysis & Problem-solving |
| 3.Inferential Comprehension | 5.Evaluate | 5.Evaluation & Synthesis (Creation) |
| 4. Evaluation | 6.Create |  |
| 5.Appreciation |  |  |

Reading comprehension taxonomies, such as Barrett's, are designed specifically to classify questions assessing comprehension of written material. More general question taxonomies, such as the Revised Bloom's Taxonomy, are designed to cover 'far more than reading comprehension' (Champeau De Lopez et al., 1997), and are more useful where evaluators have to assess a wider variety of cognitive processes.

Because of the text-based nature of South Africa's Grade 12 Home Languages examination papers, one would expect to find a more comfortable 'fit' with Barrett's Taxonomy of reading comprehension skills than the Revised Bloom's, especially in Paper 1: Language in context, and Paper 2: Literature.

As discussed in Chapter Three, a limitation of Barrett's Taxonomy is that it is not designed to evaluate either questions that test grammar skills or tasks which require students to create or produce their own text or other original product. A solution is to combine categories from the two educational taxonomies so that they can complement the weaknesses and strengths of each.

## 4.3.2 A 'combined' taxonomy for assessing Home Language examination papers

It makes sense for Umalusi and the Department of Basic Education to draw on the Revised Bloom's Taxonomy and Barrett's Taxonomy rather than to introduce a completely 'new' taxonomy of cognitive processes for a 'cognitive challenge' framework.

Table 4.2 below provides an illustrative example of how the integration of the two taxonomies could be achieved.

**Table 4.2: Typology of cognitive demand for Home Language examination analysis (based on Revised Bloom's and Barrett's Taxonomy)**

| Level of cognitive demand | Type of cognitive demand | Explanation of categorisation. Question which require students: |
|---|---|---|
| Lower order processes | **1. Recognize or recall** | To locate, identify and retrieve any kind of **explicitly stated** information, ideas, facts or details in reading material provided, or from memory of previously learned or read material (for example, names of places), and recognition of the relevance of the information, ideas, facts or details in relation to the question |
| | **2. Apply or reorganize** | To use or **apply a basic procedure** (for example, a basic grammatical rule), **to replicate a model or version** (for example, a basic visual representation, a report, memo, invitation in a highly scaffolded way where students have to recreate rather than create), or to **reorganize explicitly stated information**, ideas, facts or details from reading material or from memory of previously learned or read material **in a different way or form from what was presented** (for example, to sort, classify, match, categorize, compare, contrast, summarise or paraphrase, or consolidate **explicitly stated** information, ideas, facts or details) |

| Level of cognitive demand | Type of cognitive demand | Explanation of categorisation. Question which require students: |
|---|---|---|
| **Medium order processes** | 3.Infer, interpret or analyse | To engage in **more abstract (inferential) reasoning and interpretation**, and use conjecture, background knowledge, clues or implicit information, ideas, facts or details in reading material provided or from memory of previously learned or read material as a basis of forming hypotheses, interpreting, inferring or analysing details, relationships or ideas (for example, the significance of a theme, the motivation or nature of a character) which are **not explicitly stated** in reading or other source material |
| **Higher order processes** | 4. Evaluate or appreciate | To make critical **judgment** (for example, on qualities of accuracy, consistency, acceptability, desirability, worth or probability) using criteria provided by other sources or authorities, or students' own values, experiences, or background knowledge of the subject<br>To show emotional and aesthetic or literary sensitivity or a **reaction to the worth of psychological and artistic elements** of reading material (including literary techniques, language, forms, styles, and structuring). (For example, commenting on the effectiveness of a poetic device or image) |
|  | 5. Synthesise or create | To **integrate** ideas and information and relate parts of material, ideas, or information to one another and to an overall structure or purpose in a way that is **relational and coherent**<br>To engage in original creative thought and design and put elements together to form a coherent whole and make a new or **unique product showing emotional, aesthetic or literary sensitivity** |

The categories 'Recognize or recall' and 'Apply and reorganize' on this taxonomy form lower order processes; 'Infer, interpret or analyse' form medium order processes; and 'Evaluate or appreciate' and 'Synthesise or create' form higher order processes.

Any taxonomy can serve as no more than a 'rough guide'. Although perfect agreement cannot always be expected for categorizing questions into each of the specific categories of cognitive demand on the taxonomy, ultimately what is important is that evaluators are able to discriminate questions more broadly as requiring lower, medium, and higher level cognitive skills.

A key question raised by this research is:

**Can the assumption be made that 'alignment' with the Department of Basic Education's allocated proportions of questions at lower, medium and higher order cognitive process levels, using the above or any other educational taxonomy, mean that 40% of the examination questions are relatively easy for the average Grade 12 student to answer; 40 % of the questions are moderately challenging for the average Grade 12 student to answer; and 20% of the questions are difficult for the average Grade 12 student to answer and allow for A-grade students to be discriminated from students performing at lower levels?**

Umalusi's finer categorization of three different difficulty levels for each cognitive process category indicate that questions categorized at higher cognitive levels in both Barrett's Taxonomy and the Revised Bloom's Taxonomies are not necessarily more difficult than other questions categorized at lower process levels. Other researchers similarly argue that cognitive levels on their own do not necessarily distinguish between degrees of difficulty of examination questions for students (Guttman, 1953; Sax, Eilenberg, and Klockars, 1972; Seddon, 1978).

As discussed in Chapters Two and Three, there is also a danger that Umalusi's finer categorization of three different difficulty levels for each cognitive process category are used in a mechanical manner making judgments overly procedural, thereby constraining the 'expert' judgment of evaluators.

Findings suggest that, in order to capture the complexity of the judgment task, it may be necessary to combine relatively low-inference ratings of 'type of cognitive demand' for specific questions with more implicit expert judgments of the difficulty of examination questions.

Section 4.4 elaborates on other factors that appear to play a role in the cognitive challenge of examination questions.

# 4.4 Determining the most suitable approach to assessing the degree of difficulty of Home Language examination questions

Arguments and concerns raised in the two previous chapters suggest that there are other factors that affect the difficulty level of questions in Home Language papers that might serve as more useful indicators of cognitive challenge than degrees of difficulty as attributes of cognitive processes.

These factors relate cognitive challenge to grade level expectations and include:

1. the difficulty level of questions for Grade 12 students;

2. the degree of challenge in the texts that the Grade 12 students are required to read; and

3. the writing requirements of responses.

## 4.4.1 The difficulty level of questions for Grade 12 students

The same 'type of cognitive demand' can be made in questions in different grades, but the same question can be easier or more difficult depending on the grade level of the students who are expected to answer them.

For example, the same question requiring inferential reasoning can be far more difficult for a Grade 6 student than for a Grade 12 student. In other words, the difficulty level of the same question can change depending on the different grade levels. In assessing the difficulty level of a question, what needs to be taken into account (besides the hierarchical level of the cognitive demand), is the conceptual demand that a particular question makes on the cognitive schema of a typical student at that grade level.

Cognitive demand expectations change depending on the grade levels. What is difficult to specify, is what makes a particular 'evaluation' question difficult for a Grade 12 student when the same question may be easy for a (Home) Language university graduate or academic, for example.

Rather than linking cognitive challenge to degree of difficulty as attributes of cognitive processes (as Umalusi's instrument does), or to the explicit progression reflected in different grade level curriculum expectations (as Reeves [2005] does), it may be more constructive simply to ask Umalusi's evaluators to use their expertise in the subject, and experience of teaching it, to assess whether questions present 'cognitive challenge' for students at the Grade 12 level, and whether they are set at levels that are easy or more difficulty for typical Grade 12 students.

Evaluators could assess the difficulty level of questions as reflected Table 4.3  below:

**Table 4.3: Assessing the level of difficulty of Home Language examination questions**

| The difficulty level of questions. | | |
|---|---|---|
| This question is assessed as: | | |
| 1 | 2 | 3 |
| Easy for the average Grade 12 student to answer. | Moderately challenging but accessible for the average Grade 12 student to answer. | Difficult for the average Grade 12 student to answer. The skills and knowledge required to answer the question allows for A-grade students (extremely high-achieving students) to be discriminated from students performing at average or low levels. |

For all subjects, predictability is also a major determinant of difficulty levels of questions (for example, when the same set works are prescribed each year, or the same questions regularly appear in examination papers).

The fact that Umalusi's evaluators assess examination papers over a number of years should help them in assessing levels of predictability of examination questions. Protocols thus need to stress that evaluators must constantly gauge predictability when assessing the difficulty level of questions.

The protocol could also ask evaluators to justify their assessment of a question as 'Difficult' by explaining: *Why do you think this particular question is difficult for the average Grade 12 student? What are the characteristics that might make it difficult?*

**The proposal is that**

1. **the cognitive challenge framework includes cognitive demand ratings, as well as more inferential expert judgment of the difficulty level of questions for Grade 12 students.**

2. **the two processes are used to validate each other.**

3. **evaluator's views about the difficulty levels of each question for Grade 12 students are later compared with item analyses after students have written the examination.**

However, as discussed in Chapter Three, the degree of difficulty of the text that has to be read also has a material influence on the level of cognitive challenge – regardless of the type of questions asked about the text.

## 4.4.2 The degree of challenge in the texts that students are required to read

Although the degree of difficulty in examinations is usually revealed most clearly through the questions, indications are that a cognitive challenge framework for Home Language examinations also needs to ask evaluators to use their expertise and experience to make judgments about the degree of challenge in the (written or graphic) texts that students are required to read.

**Evaluators need to assess whether texts set and reading passages or other source material used are challenging for the grade level, and whether texts or material make high (or low) reading demands on students at the grade level.**

Research reported by DuBay (2004) shows that the best predictors of textual difficulty were two aspects of style, namely,

1. **semantic content** (for example, vocabulary and whether words used are typically outside the reading vocabulary of students; whether words are made accessible by using semantic, syntactic or graphophonic cues); and

2. **syntactic or organizational structure** (for example, sentence structure and length; whether students are likely to be familiar with the structure, for example, from reading newspapers or magazines, etc.).

3. Other aspects to take into consideration are **literary techniques** (for example, abstractness of ideas and imagery) and background knowledge required (for example, to make sense of allusions).

Table 4.4 provides an example of how this aspect could be framed.

**Table 4.4: Assessing the level of difficulty of reading requirements**

| Content, vocabulary, sentence and organizational structure, register, diction, literary techniques, abstractness of ideas and imagery, and/or background knowledge required for **reading the selection, visual text or source material** (including prescribed novels, poems, and drama) is assessed as: | | |
|---|---|---|
| 1 | 2 | 3 |
| Simple/easy for the average Grade 12 student to read | Moderately challenging but accessible for the average Grade 12 student to read | Difficult/complex for the average Grade 12 student to read |

Finally, most taxonomies make no reference to the writing demands encompassed in the response required, or the difficulty level of the text or material that students are supposed to produce. For example, a short answer requiring writing a one word answer, a phrase, or a simple sentence is easier to write than responses that require more complex sentences, a paragraph or a full essay or composition.

### 4.4.3 Writing requirements of responses

**A cognitive challenge framework may also need to reflect an interaction between the level of writing demands and the level of cognitive demands.**

For example, a question that requires low levels of cognitive demand such as recalling and reconstructing an ordered sequence of events could entail writing a few sentences, or a whole essay.

Table 4.5 provides an example of levels of difficulty for assessing 'writing' demands.

**Table 4.5: Assessing the level of difficulty of written text or representations required from students**

| 1 | 2 | 3 |
|---|---|---|
| Response requires writing a short answer (one word answer, a phrase, or sentences but **not** a whole paragraph), OR creating a concrete representation | Response requires writing one paragraph, OR creating a simple abstract (symbolic or graphic) representation | Response requires writing more than one paragraph, for example, an essay, OR creating a complex abstract (symbolic or graphic) representation |

As Umalusi integrates the number of marks allocated to each question in its analysis, this dimension of 'cognitive challenge' may be adequately covered in this way. The number of marks allocated to a question generally provides an indication of the amount of writing required.

## 4.5 Conclusion – A 'Cognitive Challenge' framework for Home Language examinations

This report has entailed evaluating different taxonomies of cognitive demand in terms of what they offer and where they have limitations.

Umalusi has attempted to address some of these challenges by offering evaluators wider options in terms of 'Easy', 'Moderate' or 'Difficult' degrees of difficulty for each 'type of cognitive demand'. These attempts have overcome some of the limitations, but they have also created new challenges. The main problem being that, where degree of difficulty cannot be clearly distinguished in terms of examples of each level, they complicate rather than simplify the task of evaluators.

The finer categorization of three different degrees of difficulty for each cognitive process category has led to an overkill of procedure of 'tick box' options, which makes it look as if rational choices can easily be made. In reality, the procedure obscures the fact that these choices are not an easy call for evaluators.

In the researcher's view, to adequately capture the complexity of the judgment task, it is necessary to combine relatively low-inference ratings of 'type of cognitive demand' for specific questions with more implicit expert judgments of the difficulty of examination questions for the average Grade 12 student.

Umalusi's evaluators are not novices in their respective fields. Along with checklist categories of cognitive demand, they should be allowed to use their knowledge of the Home Language, and experience of teaching it, to exercise their expert judgment.

Hence the proposal is to

a) reintroduce simple categories of judgment for 'type of cognitive demand', which combine the strengths of Barrett's and the Revised Bloom's Taxonomies;

b) include a more inferential expert judgment of the difficulty level of each question for Grade 12 students;

c) incorporate judgment of the degree of challenge of the written or graphic texts that students at the Grade 12 level are required to read (prescribed texts, reading passages or other source material) ;

d) take into account the writing demands of responses to questions (for example, writing a one word answer, a phrase, a simple sentence, more complex sentences, a paragraph, or a full essay or composition).


The main focus of Part 1 of this research has been on the use and value of taxonomies of cognitive demand for analysing the cognitive challenge of NSC Home Language examination questions. In Part 2, we identify a framework for judging degree of difficulty of examination questions which evaluators could use to make tacit expert notions of question difficulty or 'easiness' more explicit and transparent. Part 2 of the research also provides the new instrument for Umalusi's comparative analysis of Home Language examinations.

# PART 2:  JUDGING QUESTION DIFFICULTY

## Chapter 1: Context, Rationale, Scope and Aim

## 1.1  Introduction and objective

This report forms the second part of Umalusi's research into improving its framework for assessing and comparing the standards of the National Senior Certificate (NSC) commonly known as the 'Matric' (i.e. the qualification received on graduating from high school in Grade 12) in the different subjects and home languages across different years.

The specific focus of Part 2 of the research is on approaches to describing and classifying varying degrees of challenge of the school-leaving Senior Certificate Home Language examination questions, and the complexity of judging question difficulty prior to examination writing.

## 1.2  Context and rationale

### 1.2.1 Standards setting

To differentiate a large number of examination candidates into graded categories, examination results need to show a wide spread of marks rather than a skewed mark distribution. If all students achieve an A or B grade in an examination, results could be interpreted as reflecting high ability or proficiency on the part of all candidates (i.e. implying a homogenous group). As it is highly unlikely in large scale testing such as South Africa's school exit examination, that all candidates are 'above average', such results are more commonly an indication that the examination did not measure the entire range of levels of challenge, and the proportion of very difficult questions was insufficient.

Correspondingly, a high failure rate in large scale examinations such as the NSC could be interpreted as reflecting low ability or proficiency on the part of most candidates. As candidates usually possess a range of abilities or proficiencies, a high failure rate is more likely to indicate that an examination contained too many difficult questions (Nicholls & Miller, 1983).

To accomplish the goal of discriminating between very high achievers, those performing very poorly, and all learners in between, examiners need to vary the challenge of examination questions or tasks. They need to ensure that there are items or questions suitable for discriminating students in the different 'ability', 'proficiency' or 'competency' ranges. In other words, a range of levels of challenge needs to be evident in papers.

There also needs to be sufficient questions or items to allow for A-grade candidates to be discriminated from other high ability/proficiency candidates (i.e. questions that discriminate well amongst the higher ability candidates), and for minimally competent/proficient candidates, who should pass, to be discriminated from those candidates who should fail (i.e. questions that discriminate well amongst the lower ability candidates). Hence the relative proportions of examination questions at particular levels of challenge are important in setting examinations.

*Post hoc* standard setting methods (such as Item Response Theory/IRT) using performance data make it easier to estimate the extent to which a candidate's ability to answer a particular question correctly reflects high ability/proficiency or, alternatively, question ease, or whether a candidate's inability to answer a particular question correctly reflects a high degree of difficulty or low ability/proficiency (Nicholls & Miller, 1983:951). In other words, empirical analysis of results can be used to endorse judgement of a question's difficulty level. Unlike statistical analysis of student performance on examination questions or items 'ex *post facto*' (after writing), procedures for making judgements about the difficulty of examination questions beforehand generally have to rely on more 'subjective' human judgement, with levels of challenge most commonly being determined by a panel of subject matter experts and other experts.

## 1.2.2 Pre-examination evaluation of examination standards

In *pre-examination* evaluation of standards (where performance data are unavailable), moderators or evaluators use their knowledge of the subject, their experience in teaching it, and/or of marking examination scripts, to exercise expert judgements about whether each examination question makes low, medium or high demands on the population of examination candidates. By implication, pre-examination moderation and evaluation of levels of examination question difficulty, or 'judged difficulty', can never be an exact science.

Tools recruited for evaluating the levels of cognitive challenge of examination questions commonly include taxonomies of cognitive demand such as Bloom's Taxonomy, the Revised Bloom's Taxonomy, and Barrett's Taxonomy of reading comprehension. Bloom's Taxonomy, and others like it, identify levels of cognitive processes in a hierarchy, and are used to identify levels of cognitive processes from the highest to the lowest level of complexity.

The Revised Bloom's Taxonomy, for example, comprises the following levels of cognitive demand: Remember; Understand; and Apply (comprising lower order processes); and Analyse; Evaluate and Create (comprising higher order processes). The different levels of cognitive demand are used to classify questions in terms of increasing complexity.

For example, evaluators use the taxonomies to distinguish 'lower order questions' such as 'recognition' or 'recall' questions that require students to extract and reproduce their responses directly from a reading passage, or to restate a fact from memory, from 'higher order questions' such as 'analysis' questions that require students to break down material into its component parts.

The assumption is often made that a range of questions at lower, medium and higher order cognitive demand levels means that specified proportions of questions at each level are relatively easy for candidates to answer, moderately challenging for candidates to answer, and difficult for candidates to answer, and allow for A-grade students to be discriminated from students performing at lower levels.

## 1.2.3 Setting and monitoring National Senior Certificate standards

For South Africa's school-exit National Senior Certificate, the requirements are that students take four compulsory subjects, namely, two official South African languages (a Home Language; and a First Additional Language), Mathematics or Mathematical Literacy, and Life Orientation. In addition to these subjects, they must also take at least three other subjects from an approved list of subjects. Beyond meeting minimum requirements for simply achieving a pass in the NSC, candidates can achieve different levels of passes.

If a candidate passes the NSC with entry to Higher Certificate study, s/he has met the minimum entrance requirements in particular subjects to proceed to higher certificate courses offered by various Further Education and Training (FET) Colleges. If a candidate passes with entry to Diploma study, s/he has met the minimum entrance requirements in particular subjects to proceed to diploma courses offered by FET Colleges and Universities of Technology. If a candidate passes with entry to Bachelor's studies, s/he has met the minimum entrance requirements in particular subjects to proceed to Bachelor degree studies offered by universities and some Universities of Technology. Minimum requirements for achieving entry to Bachelor's studies are a pass in four subjects from a designated list at 50 per cent, and the remaining subjects at 30 per cent, provided that the home language is passed at 40 per cent, and the language of learning and teaching at 30 per cent.

Assessment standards and examination guidelines for the NSC prescribe the content, format and structure of the various subject examination papers (for example, the number of sections, length, time and mark allocation). They also specify the relative proportions of questions that should reflect various levels of cognitive demand. Examiners set examination papers to comply with these subject assessment and examination guidelines using taxonomies of cognitive demand and proportions provided by the Department of Basic Education.

Whereas in the past, students could take subjects at the higher or standard grade, NSC students now take all subjects on the same level which means that a single set of papers in each subject is expected to discriminate between high-achieving students, extremely high-achieving students, students performing at low levels, and students performing at levels in between. In theory, sufficient marks should be allocated to questions in examination papers that are relatively easy for candidates to answer but which students should *at a minimum* be able to answer correctly in order to pass the subject.

There should be sufficient marks allocated to questions that are moderately challenging for candidates to answer indicating that students have minimal competency/proficiency in the particular subject for entry to FET college certificate courses, or for entry to higher education diploma study. There should be a proportion of marks allocated to questions that are difficult for candidates to answer indicating minimal competency/proficiency in the subject for Bachelor's degree study at university. A portion of marks also needs to be allocated to questions which allow for very high ability or smart students to be discriminated from other high ability/proficiency candidates.

Umalusi is mandated to monitor educational standards in General and Further Education and Training (GEFT) in South Africa. In line with this mission, Umalusi has conducted several research studies that have investigated the standards of the NSC examinations. With the assistance of various subject, curriculum and assessment experts, Umalusi has developed a set of standard setting instruments and tools to analyse the cognitive challenge of examinations for various NSC subjects before students write them, and make judgements about consistency in standards in the various NSC subjects and home language examinations compared to previous years.

The chief aim of these investigations to date has been to determine whether the levels of cognitive challenge of the examinations in question were comparable across all languages and years.

## 1.2.4 Umalusi's tools and processes for evaluating examination standards

Tools recruited for distinguishing the levels of cognitive challenge of examination questions

in Umalusi's instruments include a taxonomy of cognitive demand based on the Revised Bloom's Taxonomy (Anderson & Krathwohl 2001; Anderson 2005). Panels of evaluators comprising experienced teachers, subject experts and higher education experts are provided with examination papers and marking memoranda of the various NSC subjects for the years in question. Evaluators use the relevant curricula together with Umalusi's evaluation tools to assess whether papers comply with the Department of Basic Education's examination guidelines, and check whether assessment standards as set out in the curriculum documents are covered.

Subject teams also compare different examination papers and make judgements about the relative standards over a number of years. Teams focus on rating the demands of individual questions and their related mark allocations.

Over time, Umalusi's teams of subject evaluators have adapted the 'generic' cognitive demand tool originally provided by Umalusi to make it more appropriate for use in the particular subject they are evaluating. There has also been considerable debate amongst the different subject teams about what actually constitutes the cognitive challenge of examination questions.

In line with other research (see, for example, Guttman, 1953; Sax, Eilenberg, and Klockars, 1972; Seddon, 1978), Umalusi's research processes have shown that the use of taxonomies of cognitive demand on their own do not necessarily distinguish between degrees of difficulty of questions.

For example, although questions that assess higher order processes, such as analysis, are *usually* more difficult than questions that assess lower order processes, such as recognition and factual recall (simple memory questions), questions that align to a particular type of cognitive demand are not always identifiable as being as difficult as other questions aligned to the same cognitive demand.

In other words, research has shown that various factors, besides the type of cognitive demand, contribute to the degree of difficulty of examination questions for candidates, and that questions that are difficult can make demands on students in different ways. For example:

- a recall question that asks students to recall an abstract theory, or complex content is *usually* much more difficult to accomplish than one which asks candidates to recall a simple fact (i.e. differences in content complexity). Researchers have shown that questions that require knowledge of more complex subject content are generally more difficult than questions involving knowledge of simpler subject content (Stodelsky, 1988; Soden & Pithers, 2001).

- a question that requires recalling an ordered sequence of events and which entails writing a few sentences is generally easier than one which involves the same type of cognitive demand but entails writing a whole essay (i.e. differences in 'writing' difficulty).

- literal comprehension of a reading passage comprising a contemporary magazine article is usually easier than literal comprehension of a classical work such as one of Shakespeare's plays because of differences in the content, vocabulary, sentence and organizational structure, register, diction, literary techniques, abstraction of ideas and imagery, and background knowledge required (i.e. differences in 'reading' difficulty).

In an attempt to allow evaluators to distinguish between questions which are categorized at the same cognitive demand level, but which are not of comparable degrees of difficulty,

Umalusi's instrument for evaluating the difficulty of Home Language examinations was modified to reflect the fact that each category of cognitive processes can be tested at different levels.

## 1.2.4.1 Umalusi's Home Language instrument

Umalusi's modified Home Language instrument presented the option of three different degrees of difficulty ('Easy', 'Moderate', and 'Difficult') within each category of cognitive demand. Evaluators were required first to decide on the *type of cognitive demand* required to answer a particular examination question, and then to determine the *level of difficulty* as an attribute of the cognitive level of the examination question.

Table 1.1 provides an illustrative example of one category of cognitive demand, namely, *Evaluation and Synthesis (Creation)*, extracted from Umalusi's instrument.

**Table 1.1: Example of Umalusi's levels of difficulty for Umalusi's cognitive demand category – 'Evaluation and Synthesis (Creation)'**

| 5. Evaluation & Synthesis (Creation) <ul><li>Making judgment (evaluate), critique, and recommend by considering all material available;</li><li>Weigh possibilities and do recommendations</li><li>Construct new;</li><li>Synthesise, create or find innovative solution;</li><li>Formulate new ideas</li></ul> | Easy | Opinion, giving general critique on a fairly straightforward topic. |
| | Moderate | Substantiate an opinion. Critique statements about situations made by others. Involving synthesis, critical argument; novel or abstract contexts; create poetry/a narrative. |
| | Difficult | Generalise patterns observed in situations; working with complex problems involving insight and logic-leaps; creating new solutions to problems; redesign<br>Writing a complex review / critique.<br>Rewrite information / a story for a new context and setting |

In 2010 and 2011, a challenge arose in that the Department of Basic Education's Home Language examiners and moderators used a different taxonomy (Barrett's Taxonomy) for setting and moderating some of the Home Language examination papers from the one that Umalusi was using for evaluating the cognitive demand of examination papers (based on the Revised Bloom's Taxonomy). This difference meant that Umalusi was unable to make a decision with respect to compliance of the examinations in terms of the relative proportions of questions at particular cognitive levels as specified in the Department's *Examination Guidelines Grade 12, Literature: Paper 2* (DoE, 2009a: 23).

These guidelines stipulate that examination papers should be set in such a way that questions reflect the following proportion of questions at the various levels of cognitive demand, based on Barrett's Taxonomy of reading comprehension.

**Table 1.2: Home Language Examination Guidelines requirements in terms of cognitive levels**

| Cognitive levels | Proportions of marks |
|---|---|
| 1. Literal Comprehension | 40% of total marks |
| 2. Reorganization | |
| 3. Inferential Comprehension | 40% of total marks |
| 4. Evaluation | 20% of total marks |
| 5. Appreciation | |

The discrepancy in cognitive demand taxonomies used highlighted the need for Umalusi to review and reconsider the use and application value of both taxonomies of cognitive demand, and to re-consider Umalusi's Home Language cognitive challenge instrument overall. In 2011, research was conducted into ways in which Umalusi's instrument for assessing and comparing the standards of Grade 12 Home Language examinations could be improved.

## 1.2.5 The Home Language research findings

The Home Language investigation, *The challenge of cognitive demand (Part 1 of Developing a framework for assessing and comparing the cognitive challenge of Home Language Examinations)* confirmed the importance of ensuring that a range of cognitive demands are made on students in Grade 12 examinations. The research report endorsed the need to check that examination papers are not dominated by questions that require reproduction of basic information or replication of basic procedures, and that higher level and lower level cognitive demands are adequately represented in papers.

A further recommendation of the report was that the Revised Bloom's Taxonomy (used by Umalusi) and Barrett's Taxonomy (used by the Department of Education) be combined to complement the weaknesses and strengths of each other. The report provided a taxonomy that 'married' the two taxonomies.

The Home Language report also emphasized that individual questions cannot be evaluated in isolation from other aspects of the item as a whole such as the information or 'texts' (passages, diagrams, tables, graphs, pictures, cartoons, etc.) that accompany each question, or the written or graphic texts that students are required to create or write in order to respond. The report pointed out that differences in the degree of challenge in prescribed set works can make an immense difference to the cognitive challenge of the different Home Language Literature examinations and the demands of specific questions.

The report also drew attention to the fact that a particular 'evaluation' question (for example) may be more difficult for a typical Grade 12 student than for a university graduate. A question requiring inferential reasoning may be far more difficult for the average Grade 6 student than for the average Grade 12 student. As questions must be of appropriate difficulty for the students to whom they are administered, difficulty is thus also, to a certain extent, relative to the particular reference group.

Thus, Part 1 of the Home Language report recommended that, in addition to identifying the level of cognitive demand made in each examination question, instruments for assessing and comparing the NSC Home Languages examinations incorporate judgement of

- the degree of challenge of the written and graphic texts that students are required to write for examinations may differ significantly in terms of their intrinsic difficulty, making the examination more (or less) cognitively challenging. For example, a short answer requiring writing one word, a phrase, or a simple sentence is usually easier to write than a response that requires more complex sentences, a paragraph or a full essay or composition.

- the level of challenge of the written and graphic texts that candidates have to read. Texts (reading passages or other source material) selected may differ significantly throughout different years, and in different examinations, in terms of their intrinsic difficulty, making the examination more (or less) cognitively challenging.

- the level of the academic demands that a question makes on students at a particular grade level in terms of subject matter knowledge and changes in content and cognitive expectations across different grade levels.

The research pointed to problems with the conception of *degree of difficulty* as an attribute of the level of cognitive demand of examination questions, and with judgement strategies involving the use of 'tick box' options for grading the difficulty level of examination questions or items. It found that, although the modifications to the Home Language instrument had helped to overcome some of the limitations of using cognitive demand taxonomies on their own, they had created new challenges. The finer categorisations of 'Easy', 'Moderate' and 'Difficult' for each level of cognitive demand made 'making judgments' quite complex and time consuming.

The process of trying to match questions to Umalusi's definitions of levels of difficulty for each cognitive demand category tended to distract evaluators from the idea that, what they were being asked to do, was to use their internalised expertise and experience to determine whether the range of questions in examinations would make it possible to discriminate between high-achieving students, students performing at low levels, and students performing at levels in between. The danger of this approach is that procedural compliance tended to take precedence over expert judgement.

The Home Language report argued, in addition, that to adequately capture the complexity of the judgment task  it might be necessary to combine relatively low-inference ratings of 'type of cognitive demand' for specific questions with a set of more nuanced expert judgment of question difficulty. A key recommendation was that evaluators be given opportunities to use their knowledge of the subject, their experience of teaching it, and/or of marking Grade 12 examination scripts, to exercise more inferential judgements about whether each examination question is 'Easy', 'Moderate' or 'Difficult' for the 'average' Grade 12 student. In other words, the report suggested that evaluators also need to consider the difficulty of examination questions with reference to a defined target candidate.

Table 1.3 below provides a brief explanation of Part 1 of the report's suggested ratings for these three levels of difficulty.

**Table 1.3: Proposed degree of difficulty of examination questions in Part 1 of Umalusi's the Home Language research report**

| This question is assessed as: | | |
|---|---|---|
| **1** | **2** | **3** |
| Easy for the average Grade 12 student to answer | Moderately challenging for the average Grade 12 student to answer | Difficult for the average Grade 12 student to answer. The skills and knowledge required to answer the question allows for A-grade students (extremely high-achieving students) to be discriminated from students performing at average or low levels |

In terms of the judgement procedures used in South Africa's school exit-examinations, the credibility of the procedures employed by standard setting and quality assurance bodies, such as Umalusi, is crucial. In high stakes examinations where public trust in the examining system is essential, a key issue is the transparency of the judgement processes. The conclusions and recommendations in Part 1 of the Home Language report raised the need for Umalusi to further explore and identify guidelines for assessing the difficulty level of examination questions and a strategy that entails rendering expert judgement more transparent.

In summary, the findings of the study highlighted that some of the challenges of comparing the degree of cognitive challenge of the eleven language examinations include (i) unguaranteed consistency with which the instrument has been applied across languages, (ii) the difficulty of making comparisons with regard to Papers 2 and 3 because of the number of the different choices that candidates could opt for, and (iii) the difficulty in applying the instrument across the three papers because Barrett's taxonomy, on which the instrument was based, was originally intended to assess comprehension alone.

In addition, the evaluation process uncovered that there are factors other than cognitive level that influence the degree of difficulty of an examination paper. One of the recommendations emanating out of the study was the need for construction of a new Home Language examination evaluations instrument.

# 1.3   Scope and purpose of this research

The main focus of Part 1 of the Home Language research report was on reviewing and evaluating taxonomies of cognitive demand used for categorizing Language examination questions, in terms of what they offered, and where they had limitations. Part 2 of the report explores a more inferential approach to assessing the difficulty level of examination questions that goes beyond a 'tick box' approach. It proposes guidelines and a new instrument and a framework for making the expert judgement process more transparent.

# 1.4   Research approach

The research approach for Part 2 involved:

1.   Observation of Umalusi's 2011 teams of Home Language examination evaluators using Umalusi's Home Language instrument, '*A comparative analysis of the National Senior Certificate Home Language examination papers*', with actual examination papers. The

purpose was to incorporate an empirical dimension into the study.

2. Surveying literature on standard setting processes and item/question difficulty. The main objective was to identify possible guidelines for making expert judgement of examination questions.

3. Synthesis of available evidence from the literature review together with data and ideas generated from the 2011 Home Language examination evaluators' workshop to answer the following key question: **What framework could best serve as guidelines for assessing and comparing the difficulty level of questions in National Senior Certificate (NSC) examinations? What are key categories and concepts for consideration when judging questions difficulty?**

4. Using the above synthesis to draft a new version of Umulusi's instrument '*A comparative analysis of the National Senior Certificate Home Language examination papers*'.

5. Participation in the piloting of the revised the instrument with Umalusi's 2012 teams leaders of Home Language examination evaluators for the eleven official languages in June 2012.

6. Participation in a workshop (2-4 July 2012) designed to provide a platform for all Home Language examination role players to discuss and critique the proposed conceptual framework for the new instrument. The purpose of the workshop was to refine the conceptual underpinnings of the new instrument, and provide exemplar items from the 2011 examination papers for each level of difficulty and level of cognitive demand.

7. Using the June pilot and the July workshop to finalize the instrument for comparing the 2009 – 2012 Home Language examination papers.

# 1.6  Aim

The aim of Part 2 of the report is to

- identify different pre-examination approaches to describing and classifying difficulty levels of examination questions;

- compare different methods and models for judging question difficulty for candidates; and

- propose guidelines that could be used by evaluators when judging degrees of difficulty of examination questions and which evaluators could use to make tacit notions of question or item difficulty or 'easiness' explicit or public.

- present a new instrument for Umalusi's comparative analysis of Home Language examinations.

# 1.7  Outline of report

Chapter One frames the report by providing the rationale and background to the study.

Chapter Two provides an overview of different approaches to and models for judging the difficulty of examination questions.

Chapter Three identifies key features of a conceptual framework for thinking about the level of difficulty of examination questions for Grade 12 students and discusses key facets of the new instrument.

Appendix A provides the new instrument.

# Chapter 2: Judging question difficulty: Approaches and Challenges, Methods and Models

## 2.1 Introduction

This chapter considers different approaches to describing and classifying levels of difficulty of examination questions. It compares different methods and models for judging question difficulty or 'easiness' for examination candidates. The chapter incorporates consideration of the usability of different judgement methods for Umalusi's purposes and for given groups of evaluators, in particular Umalusi's subject team evaluators.

## 2.2 Approaches to describing and classifying question difficulty

Broadly speaking, two main ways of approaching pre-examination moderation and evaluation of examination question difficulty levels are evident.

One way of thinking about difficulty levels of examination questions is through a conception of a continuum of 'objective' or 'criterion-referenced' difficulty levels (Nicholls & Miller, 1983). This conception is based on specific properties or qualities – such as size or complexity – which enable evaluators to compare and 'order' tasks or questions in terms the ability they demand. An illustrative example of this conception is thinking about difficulty levels in relation to the number of pieces in jigsaw puzzles (Nicholls & Miller, 1983: 951).

Puzzles that have the most pieces are generally considered more difficult than puzzles that have the least pieces. Puzzles that have a few more pieces are usually considered more difficult than those with fewer, or less difficult than those with the most pieces. In this conception of the judgement task, task or question difficulty is judged independently of norms for a particular reference group (for example, an age or grade cohort). The idea is that higher levels of ability are required by 'objectively' defined more difficult tasks or questions.

Another way of thinking about difficulty levels is through a conception of the relative difficulty for target student/s within a particular reference group (Nicholls & Miller, 1983). This basis for judging degree of difficulty applies when questions are judged as more difficult if fewer members the target group of students in the reference group (for example, NSC candidates) should/would be able to answer questions correctly or well. Correspondingly, items or questions are judged as 'easy' if most members of the target students could/should be able to answer them correctly or do them well.

The judgement task involves deciding which questions nearly all the target candidates should/would get correct or answer well, and which questions only a few very clever or competent candidates would be able to answer correctly or well. In this conception, judgement is not based on 'objectively' defined difficulty levels but on difficulty relative to a particular target student or group of target students within a reference group.

Essentially, Part 1 of the Home Language report advocated an approach to describing and classifying degrees of difficulty of Home Language examination questions that incorporates both of conceptions of difficulty levels. The report suggested that to judge the degree of difficulty of examination questions, evaluators need to consider the intrinsic difficulty of the particular question as well as the demands that each question makes on the cognitive schema of a target Grade 12 examinee.

The proposed ratings for degrees of difficulty shown earlier in Table 1.1 in Chapter 1 of this report indicate that evaluators need to consider the difficulty of examination questions relative to the properties or qualities that make questions easy, moderately challenging, or difficult for the 'average' Grade 12 student to answer. In other words, the suggestion is that assessment of question difficulty is linked to a particular target student within the population of NSC candidates – the 'average' Grade 12 examination candidate.

Section 2.3 discusses methods and challenges related to judgement strategies linking question difficulty to a particular target student or group of target students.

## 2.3 Methods for judging question difficulty for target students

One of the most exhaustively investigated judgment methods for evaluating question difficulty for a particular target student or student group appears to be Angoff's standard setting methodology (Hein & Skaggs, 2010). In this method, evaluators identify items or questions that candidates should be able to answer in order to be declared 'minimally competent' (for example).

In the most commonly used version of Angoff's standard setting methodologies (Angoff, 1971), evaluators first need to develop a common understanding of a *hypothetical* examinee (variously referred to in the literature as 'a minimally competent'; 'borderline'[8]; or 'minimally qualified' candidate) (Hein & Skaggs, 2010). They are then instructed to envision a group of 100 such candidates and estimate the percentage or proportion (number) of this target group who would answer a given question in an examination/test correctly.

For example, evaluators are asked: *"What percentage of the target candidates would answer the question correctly? The highest rating you can have is 95% and the lowest rating is 25%. For example, an easy question might have a rating of 95% and a difficult question might have a rating of 25%."*

Because evaluators often find it difficult to conceptualize 100 target candidates and arrive at percentage estimates (Impara & Plake, 1997) various modifications of this method have been made. As the classroom unit tends to be more familiar to evaluators, one modification is for evaluators to envision a hypothetical classroom of the target students and then assess the percentage or proportion of target students in the *hypothetical class* that would answer the question correctly or well (Hein & Skagg, 2010).

For example, evaluators are asked to imagine a class made up of 'minimally competent' candidates, and then look at each examination question and answer the following questions: '*Based on what you believe a minimally competent student to be, do you think that two thirds of the students in the class will answer this question correctly? If you think that two thirds of the class of target examinees will answer the question correctly, then on the rating sheet, circle 'Yes' next to the question number. If you think that two thirds of the students will/could not answer the question correctly, then on the rating sheet, circle 'No' next to the question number.*' (Skagg & Hein, 2011)

Another modification of the methodology is to remove the estimation of the proportion or percentage of a group from the process (Hein & Skagg, 2010). Evaluators are instructed to envision just one *hypothetical* candidate (who is 'minimally competent', for example) and

---

[8] In this conception, a hypothetical student is a typical candidate who is 'at the border between two achievement levels' (Hein & Skagg, 2010: 37)

judge whether an individual hypothetical examinee would/could or would/could not answer a particular question correctly (yes–no dichotomously scored as 0 – 1).

For example, evaluators are asked: "*Would a minimally competent candidate answer the question correctly*?" A score of one is given if the evaluator thinks the question would be answered correctly by the hypothetical student and a score of zero is given for each question that the evaluator thinks would be answered incorrectly (Skagg & Hein, 2011).

As a modification of this method, Impara and Plake (1997) instructed evaluators to think of a student with whom they are *familiar*, in other words to conceptualize *an actual student* (for example, an individual in their class) who was typical of the target group when making ratings, rather than a *hypothetical* candidate. Bearing this student in mind, they were asked to indicate whether the student would answer each question correctly (yes–no).

Of interest is that, during a debriefing session, evaluators reported that an actual person was easier to envision than a hypothetical one. In other words, it seems that a procedure that requires evaluators to consider actual students, especially students whom they have taught is easier for evaluators to undertake (Hein & Skagg, 2010). However, evaluators may experience difficulty with making judgements about individual questions, because an *actual* individual's performance tends to vary from question to question, or from day to day (Hein & Skagg, 2010).

Researchers have also called attention to other challenges related to linking question difficulty to particular target examinee/s. In particular, attention has been drawn to difficulties associated with the degree of variability in individual evaluators' conceptions of the target student, or, of a classroom of the target students (Skagg & Hein, 2011).

## 2.3.1 Variability in evaluators' conceptions of the target students

Essentially, the student/s envisioned or 'chosen' by each evaluator may not be representative of the overall population of target students. Because evaluators need to estimate performance or ability of an entire target population, what is difficult to gauge, is whether individual evaluator's conceptions of the target students as a collective is representative of the population of target students in general (Hein & Skagg, 2010).

For example, in Umalusi's case, if evaluators are asked to rate the level of the challenge of a particular question or item for the 'average' Grade 12 candidate, it is very important that evaluators' conception of the 'average' student is representative of the entire population of average Grade 12 candidates for all schools in the country. A danger is that, if they reference only students whom they have taught, or of whom they have direct experience, evaluators may reference students from only one area or region in the country, or from only a certain language group or group of schools or classrooms (Hein & Skagg, 2010).

Much depends on the sample of evaluators and their experience and exposure to the population of candidates in general (Hein & Skagg, 2010). If all or almost all evaluators are representative of one racial or language group, or from one geographic region in the country, they may not have adequate knowledge and experience of students who are representative of Grade 12 candidates in the country as a whole. Using evaluators of different racial and language compositions, or from other regions, could result in different difficulty level ratings. The implication is that panel members need to reflect ethnic and geographic diversity.

This aspect could present a challenge in relation to panels of evaluators for some of South Africa's Home Language examination papers in minority languages. It might be difficult to find evaluators with relevant expertise who are not representative of one language group from one geographic region in the country and whose conception of the 'average' student is representative of the population of average home language candidates for all schools in the country.

Section 2.4 considers different models for making judgments of degrees of difficulty of examination and proposes a strategy for making an inferential approach more explicit.

# 2.4 Models for judging question difficulty for candidates

Cresswell (1997 in Crisp, 2010:22) identified four models that are used for *grading students' responses*. The models, which are explained below, appear to be relevant to, and have application value for, judging examination question difficulty for candidates.

1. A judgement model where an evaluator identifies the presence or absence of specified properties and employs set rules to compute whether a sufficient number of combined properties are present. With regard to judging examination question difficulty, this model implies setting up explicit pre-determined criteria for grading questions that must be used to determine and 'compute' the degree of difficulty.

2. A judgement model which is similar to the above model in that a cognitive schema is consciously used, but the model differs from the above model in that it does not incorporate prescribed rules of combination and computation. The evaluator identifies the presence or absence of specified properties and combines these features through a more flexible, less computational process to make a judgement. In relation to judging examination question difficulty, this model implies the selection of those properties or qualities from provided criteria that are relevant to the question under consideration. The selected criteria may even be slightly modified.

3. A model where the judgement process is not broken down into criteria that should be used to make judgements. Rather, the model relies on the evaluator's first and immediate perception. With regard to judging examination question difficulty, this model implies the use of 'tacit internalized notions' of the degree of difficulty without necessarily providing justification. In other words, the model suggests the use of a 'gut feeling or instinct' (Crisp, 2010: 21) (for example, for identifying a difficult question that distinguishes an A-Grade student) which is developed through years of relevant teaching, examining and/or marking experience. What makes some questions inherently more difficult than others, is not directly or explicitly addressed.

4. A model where the evaluator's first impression is iteratively revised until the judgement stabilises. This model implies the use of tacit internalized notions of difficulty levels, but also suggests that decisions need to be defensible and that the qualities or properties in the question that informed the judgement need to be identified, provided and defended.

## *2.4.1 A model for Umalusi's Home Language evaluation*

Umalusi's Home Language report (Part 1 of *Developing a framework for assessing and comparing the cognitive challenge of Home Language Examinations*) argues that a rigid judgement strategy for assessing the difficulty of examination questions involving the use of prescribed criteria runs the risk of evaluators losing a sense of judgement as a coherent whole. The report maintains that expert judgment of difficulty levels of examination questions or items usually involves using multiple and interdependent criteria. The complexity of

the judgement task lies in the evaluators' ability to recognize subtle interactions and links between different aspects of each question's difficulty and make holistic judgements. Such judgements are complicated, non-linear processes. The challenge of evaluating question difficulty is that the judgements involve a relational logic.

A corresponding challenge is that the tacit nature of this expert knowledge makes it difficult to make public the notions of question difficulty (Fisher-Hoch & Hughes, 1996). If Umalusi's evaluation process is to gain public confidence, evaluators need to make their judgments transparent. Paradoxically, what makes the judgement process more or less transparent, is the extent to which what makes one question more difficult than another for candidates is explicitly addressed. To justify their decisions and provide an account of how these decisions are made, evaluators need to be able to identify where the difficulty or ease in each question resides.

Essentially what evaluators need to consider and make explicit, are the features of a question that are the sources of difficulty or 'easiness' for the 'average' Grade 12 examination candidate.

The proposal that this report makes is that Umalusi combine features of models 2 and 4 above in a model where

- evaluators draw on their expert tacit internalized notions of difficulty levels to make a judgement using a conceptual framework provided to help identify and articulate the properties or qualities that are relevant to each question under consideration;

- space is provided for individuals to discuss and compare their initial judgements or opinions of difficulty levels with those held by other panel members;

- team members revise their initial decisions until a convergence of decisions is reached; and

- decisions are justified and the qualities or properties in the question that informed the judgement are articulated and provided.

One danger of such a process which includes decision-making involving consensus in awarding difficulty ratings is that evaluators who are not confident about their views can be unduly influenced by the status or personalities of other more influential members of the subject panel (see Murphy et al. in Crisp, 2010:24; and Buckendahl, Smith, Impara & Plake, 2002). Rather than trusting their independent judgements, less confident team members may simply revise their difficulty level ratings for each question to match those of the other more dominant members. Hence the proposal that each evaluator first works on his/her own, and then works together in his/her subject/language teams sharing and discussing views.

The tacit nature of the kind of expert knowledge needed can make it difficult for evaluators to articulate their understandings and generate a common understanding of what constitutes levels of difficulty. Thus it seems that what is needed for Umalusi's purposes is a framework which helps to make explicit what is tacit, but which is not written in a prescriptive way that prevents evaluators from grappling with nuances and connections.

The most realistic approach appears to be for evaluators to have a simple frame of reference for thinking about item or question difficulty as a means of guiding and supporting their judgement of degree of difficulty. Ideally, guidelines for assessing the difficulty level of examination questions or items need to act as a heuristic device which helps to shape and guide rather than prescribe, or dictate and control, the judgement process.

Chapter 3 provides such guidelines that evaluators could use for judging levels of challenge of examination questions.

# Chapter 3:  Guidelines for judging question difficulty

## 3.1  Introduction

This chapter offers a conceptual framework for thinking about question difficulty. The framework is not intended to be prescriptive. Rather the guidance provided in the framework on where the challenge in a question might reside, is intended to provide support for evaluators in making complex decisions about what makes a particular question or item easy, moderately challenging, difficult or very difficult for the average Grade 12 examination candidate.

The framework is intended to influence evaluators' perceptions of question difficulty and to help them develop a more explicit understanding of what aspects make a question difficult.

## 3.2  A Framework for thinking about question difficulty

The four-category conceptual framework for thinking about item or question difficulty provided by Leong (2006) of the *Singapore Examinations and Assessment Board* seems useful for Umalusi's purpose. In developing the framework Leong drew on work of other researchers such as Pollit et al. (1985)[9] and Osterlind (1990 in Leong, 2006) on where the sources difficulty (SODs) or easiness (SOEs) in a test item or examination question might reside.

A number of other researchers have developed similar tools for gauging cognitive challenge. For example, Edward and Dall'Alba (1981) developed a scale for analysing the cognitive challenge of secondary school science material and lessons in Australia. Their scale also comprises four dimensions: *Complexity* defined as 'the complexity of each component operation or idea and the links between them' (Hughes, Pollit & Ahmed, 1998:5); Openness, defined as 'the degree to which a task relies on the generation of ideas' (Edward and Dall'Alba, 1981:2), *Implicitness*, defined as 'the extent to which the learner is required to go beyond data available to the senses' (Edward & Dall'Alba, 1981:3); and *Level of Abstraction* defined as 'the extent to which a task deals with ideas rather than concrete objects or phenomena' (Hughes, Pollit & Ahmed, 1998:3).

Hughes, Pollit and Ahmed (1998) refined and modified the Edward's scale and drew on Kelly's Repertory Grid technique (1955) to develop the 'CRAS' scale for identifying and quantifying the demands and difficulties of History, Chemistry and Geography examination questions. The four dimensions of the 'CRAS' scale are: *Complexity; Resources* defined as 'the use of data and information' (Hughes, Pollit & Ahmed, 1998:5); Abstractness; and Strategy defined as 'the extent to which the student devises (or selects) and maintains a strategy for tackling and answering the question' (Hughes, Pollit & Ahmed, 1998:5).

The rationale for selecting Leong's framework over other tools is the belief that Umalusi's teams of Home Language evaluators along with other stakeholders will find the framework easier to understand. Other tools that have been developed appear to be more complicated to use and not to have as much application value for text-based Language examination papers.

---

[9] Pollit et al., (1985) used empirical analysis of common errors made by students when answering difficult questions to identify SODs and SOEs.

Briefly, Leong's adapted framework comprises the four general categories of difficulty:

- Content (subject/concept) difficulty;

- Stimulus (question) difficulty;

- Task (process) difficulty; and

- Expected response difficulty.[10]

As will be shown later, the different categories play a role in explaining certain aspects of question difficulty depending on how they are embodied in particular questions. Although it is often the case that question difficulty is located within more than one of the above categories of difficulty at a time, examples of Home Language examination questions displaying elements the four categories are also provided.

For each general category Leong (2006) draws on the literature to make a further distinction between 'valid' and 'invalid' sources of question difficulty or easiness. 'Valid difficulty' or 'easiness' have their source in the requirements of the question and are **_intended_** by the examiner (Ahmed & Pollit, 1999).

Invalid sources of difficulty or easiness are defined as those features of question difficulty or easiness that were **_not intended_** by the examiner. Invalid sources prevent the question from assessing what the examiner intended and are likely to prevent candidates from demonstrating their true ability or competence (Ahmed & Pollit, 1999, see also Hannah Fisher-Hoch & Hughes, 1996:2). They are factors irrelevant or indirect to the construct being measured.

Leong's framework includes examples of probable invalid sources of difficulty for each of the four categories of difficulty. For example, grammatical errors in a question that could cause misunderstanding are possible invalid sources of question difficulty, because the difficulty in answering the question could lie in the faulty formulation of the question, rather than in the intrinsic difficulty of the question itself (stimulus difficulty); 'students may misunderstand the question and therefore not be able to demonstrate what they know' (Ahmed & Pollit, 1999:2).

The framework outlined in section 3.2.1 – 3.2.4 draws directly on Leong's framework for thinking about question difficulty and incorporates additions and amendments in line with the findings and recommendations from Part 1 of the Home Language Report and piloting of the instrument.

In the framework:

- the demands of the reading required to answer a question forms an important element of 'Stimulus difficulty'.

- the demands of the written text or representations students are required to produce for their response form an important element of 'Task difficulty'.

- the level of cognitive demand of a question forms an element of 'Task difficulty'. In the framework, the level of cognitive demand of a question is only one attribute of the degree of difficulty of the examination question. It is just one of the characteristics that influences comparative judgements made.

---

[10] Osterlind (1990) suggests that item difficulty could be located in four different areas – content assessed; stimulus; task to be performed; and expected response difficulty.

### 3.2.1 Content or concept difficulty

Content or concept difficulty indexes the difficulty in the subject matter and/or concept/s assessed. In this judgment of the item/question, difficulty lies in the academic demands that questions make and the grade level boundaries of the various 'elements' of the knowledge domain (facts, concepts, principles and procedures associated with the subject).

For example, questions that assess 'advanced content', that is, subject knowledge that is considered to be in advance of the grade level curriculum, are *likely* to be difficult or very difficult for most candidates. Questions that assess subject knowledge which forms part of the core curriculum for the grade are *likely* to be moderately difficult for most candidates. Questions that assess 'basic content' or subject knowledge candidates would have learnt at lower grade levels, and which would be familiar to them are *unlikely* to pose too much of a challenge to most candidates. Questions that require general everyday knowledge or knowledge 'real life' experiences are *usually* easier than those that test more specialized school knowledge.

Questions involving only concrete objects or phenomena are *usually* easier than those that involve more abstract constructs or ideas. For example, questions which test students' understanding of theoretical, conceptual or de-contextualized issues or topics, rather than their knowledge of specific examples or contextualized topics or issues tend to be more difficult. Deriving and providing an abstract concept, underlying principle or generalization from a given example is *usually* more difficult than deriving and providing an example of a general statement, principle or concrete concept.

Content difficulty may also be varied by changing the number of knowledge elements assessed. *Generally* the difficulty of a question increases with the number of knowledge elements assessed. Questions that assess students on two or more knowledge elements are *usually* (but not always) more difficulty than those that assess a single knowledge element. Assessing students on a combination of knowledge elements that are seldom combined *usually* increases the degree of difficulty (Ahmed, Pollitt, Crisp, & Sweiry, 2003 in Leong, 2006:3).

Examples of 'difficult' English Home Language examination questions involving more abstract ideas (content difficulty)[11] are: Comment on the suitability of the title of the poem *The Weeping of the Penny Whistle* by Peter Horn; and Explain how humour is created in the (provided) *Madam and Eve* cartoon.

Leong provides the following examples of probable invalid sources of content difficulty of questions:

- Testing obscure or unimportant concepts or facts that are hardly mentioned in the curriculum, and which are unimportant to the curriculum learning objectives.

- Testing advanced concepts that candidates are extremely unlikely to have had opportunities to learn.

---

[11] As noted earlier, question difficulty is often located within more than one of the four categories of difficulty at a time.

## 3.2.2 Stimulus difficulty

Stimulus difficulty refers to the difficulty of the linguistic features of the question (linguistic complexity) and the challenge that candidates face when they attempt to read and understand the words and phrases in the question and in the information or 'texts' (diagrams, tables and graphs, pictures, cartoons, passages, etc.) that accompany the question.

For example, questions that contain words and phrases that require only simple and straightforward comprehension are *usually* easier than those that require the candidate to understand subject specific phraseology and terminology (e.g. idiomatic or grammatical language not usually encountered in everyday language), or that require more technical comprehension and specialised command of words and academic language (e.g. everyday words involving different meanings within the context of the subject) (Rothery, 1980 in Fisher-Hoch & Hughes, 1996)

Questions that contain information that is tailored to an expected response, that is, that contain no irrelevant information, are generally easier than those that require candidates to select relevant and appropriate information or comprehend a large amount of information for their response. Questions that depend on reading and selecting content from a text (including a graph, picture, cartoon, etc.) *can* be more challenging than questions that do not depend on *actually* reading the text, because they test reading comprehension skills as well as subject knowledge. Questions that require candidates to read a lot can be more challenging than those that require limited reading.

Although the degree of difficulty in examinations is usually revealed most clearly through the questions, evaluators also need to consider the complexity of the text and the degree of challenge in written or graphic texts that students are required to read and interpret in order to respond. They need to consider whether texts set, and reading passages or other source material used are challenging for the grade level, and whether prescribed texts or source material make high (or low) reading demands on students at the grade level.

Predictors of textual difficulty include

- semantic content – for example, if vocabulary and words used are typically outside the reading vocabulary of Grade 12 students, 'texts' (passage, cartoon, diagram, table, etc.) are *usually* more difficult. If words/images are made accessible by using semantic, syntactic or graphophonic cues, 'texts' are *generally* easier. (DuBay 2004)

- syntactic or organizational structure - for example, sentence structure and length. If, for instance, students are likely to be *familiar* with the *structure* of the 'text', for example, from reading newspapers or magazines, etc. 'texts' are *usually* easier than when the structure is unfamiliar. (DuBay 2004)

- literary techniques – for example, abstractness of ideas and imagery - and background knowledge required, for example, to make sense of allusions. For example, if the context is unfamiliar and candidates do not have access to the context which informs a text (passage, cartoon, diagram, table, etc.), they are expected to read, and which informs the question they are supposed to answer and the answer they are expected to write, then constructing a response is *likely* to be more difficult than when the context is familiar (Halliday 1973, 1978).

Another important factor in stimulus difficulty is presentation and visual appearance. For example, type face and size, use of headings etc. can aid 'readability' (Mobely, 1987 in Fisher-Hoch & Hughes, 1996).

An example of a 'difficult' English Home Language examination question because of the demands involved in understanding the question (stimulus difficulty) is: Your school is preparing a pamphlet to be given to the new Grade 8 learners. You have been asked to write, for inclusion in this pamphlet, instructions on how to manage the challenges of being in high school. Write out these instructions.

The following are examples of probable invalid sources of stimulus difficulty

- Meaning of words unclear or unknown

- Difficult or impossible to work out what the question is asking

- Questions which are ambiguous

- Grammatical errors in the question that could cause misunderstanding

- Inaccuracy or inconsistency of information or data given

- Insufficient information provided

- Unclear resource (badly drawn or printed diagram, inappropriate graph, unconventional table)

- Dense presentation (too many important points packed in a certain part of the stimulus).

## 3.2.3 Task difficulty

Task difficulty refers to the difficulty that candidates are confronted with when they try to generate or formulate an answer. For example, in most questions, to generate a response, candidates have to work through the steps of a solution. *Generally*, questions that require more steps in a solution are more difficult than those that require fewer steps. Task difficulty may also be mediated by the amount of guidance present in the question.

Although question format is not necessarily a factor and difficult questions can have a short or simple format, questions that provide guided steps are *generally* easier than those that are more open ended and require candidates to form their own response strategy, work out the steps and maintain the strategy for answering the question by themselves. A high degree of prompting (a high degree of prompted recall, for example) tends to reduce the degree of task difficulty.

Questions that test specific knowledge are *usually* less difficult that multi-step, multiple-concept questions. A question that requires the candidate to use a high level of appropriate subject, scientific or specialized terminology in their response *tends* to be more difficult than one which does not. A question requiring candidates to create a complex abstract (symbolic or graphic) representation is *usually* more challenging than a question requiring candidates to create a concrete representation.

A question requiring writing a one word answer, a phrase, or a simple sentence is *often* easier to write than responses that require more complex sentences, a paragraph or a full essay or composition. Narrative writing is usually easier than writing discursively (argumentatively or analytically).

In some subjects such as language(s) or history, where one of the goals is that students learn to express themselves well (in English, etc.) and writing skill is part of what students are supposed to learn, some questions reflect expected response difficulty simply by 'creating the space' for A-grade candidates to demonstrate genuine insight or good argumentation, and

to write succinctly and coherently about their knowledge. In contrast, questions which require continuous prose or extended writing may also be easier to answer correctly than questions that require no writing at all or single letter answer (such as multiple choice), or a brief response of one or two words or short phrase(s), because they test very specific knowledge.

Cognitive demand forms another aspect of task difficulty. Some questions test thinking ability and students' capacity to deal with ideas. Questions that assess inferential comprehension or application of knowledge, or that require students to take ideas from one context and use them in another, for example, *tend* to be more difficult than questions that assess recognition or retrieval of basic information. Tasks requiring recall of knowledge are *usually* more difficult than tasks that require simple recognition processes. When the resources for answering the question are included in the examination paper, then the task is *usually* easier than when candidates have to use and select their own internal resources (their own knowledge of the subject) to answer the question.

An example of a 'difficult' English Home Language comprehension question involving higher level reasoning (task difficulty) is:

> *Disqualified (in South Africa) from local first-class cricket on the grounds of race, D'Oliviera (Basil D'Oliviera, a world-class talent who just happened to have the 'wrong' colour of skin) went to live in England in 1960, becoming one of the stars of the English team. When he was selected for a 1968 tour of South Africa, the apartheid government barred him – an act of folly that offended even the crustiest British conservatives, and turned South Africa into an international sporting pariah.*

What does 'When he was selected … crustiest British conservatives' suggest about the difference between British and South African conservatives at the time?

An example of a discursive essay involving offering one's own opinion and reasoning (task difficulty) is: Should cellphones be used as a tool in education? Write an essay in which you clearly express your views on this topic.

The following are examples of invalid sources of task difficulty

- Level of detail required in an answer is unclear.

- Context is unrelated to or uncharacteristic of the task that candidates have to accomplish.

- Details of a context can distract candidates from recalling or using the right bits of their knowledge.

- Question is unanswerable.

- Illogical order or sequence of parts of the questions.

- Interference from a previous question.

- Insufficient space (or time) allocated for responding.

Question predictability/task familiarity. If the same questions regularly appear in examination papers or have been provided to schools as exemplars, students are likely to have had prior exposure, and practised and rehearsed answers in class (for example, when the same language set works are prescribed each year).

## 3.2.4 Expected response difficulty

Expected response difficulty refers to the difficulty imposed by examiners in a mark scheme and memorandum. Mark allocations affect the amount of information and level of answers students are expected to write. Thus this location of difficulty is more applicable to 'constructed' response questions, as opposed to 'selected' response questions (such as multiple choice, matching/true-false). For example, when examiners expect few or no details in a response, the question is generally easier than one where the mark scheme implies that a lot of detail is expected.

Another element in this category of difficulty is the complexity in structure of an expected response. When simple connections between ideas are expected in a response, the question is generally easier to answer than a question in which the significance of the relations between the parts and the whole is expected to be discussed in a response. In other words, a question in which an unstructured response is expected is *generally* easier than a question in which a relational response is expected. A response which involves combining or linking a number of complex ideas is *usually* more difficult than a response where there is no need to combine or link ideas.

A further aspect of expected response difficulty is the clarity of the allocation of marks. Questions are *generally* easier when the allocation of marks is straight-forward or logical (i.e. 3 marks for listing 3 points) than when the mark allocation is indeterminate (e.g. when candidates need all 3 points for one full mark or 20 marks for a discussion of a concept, without any indication of how much and what to write in a response). This aspect affects difficulty, because candidates who are unclear about the mark expectations in a response may not produce a sufficient amount of information in answer to the question in their response that will earn the marks that reflect their ability.

Some questions are more difficult to mark accurately than others. Questions that are harder to mark and score objectively are generally more difficult for candidates to answer than questions that require simple marking or scoring strategies on the part of markers (Suto & Nádas, 2009). For example, recognition and recall questions are *usually* easier to test and mark or score objectively, because they usually require the use of matching and/or simple scanning strategies on the part of markers.

More complex questions requiring analysis (breaking down a passage or material into its component parts, and understanding relationships between the parts), evaluation (making judgments, for example, about the worth of material or text, or about solutions to a problem), synthesis (bringing together parts or elements to form a whole), and creativity (presenting original thought) are generally harder to mark/score objectively. The best way to test for analysis, evaluation, synthesis and creativity is usually through essays. Such essays generally require the use of more cognitively demanding *marking* strategies such as interpreting and evaluating the logic of what the candidate has written.

Questions where a wide range of alternative answers or response(s) is possible also tend to be more difficult. In contrast, questions may be so open-ended that learners will get marks even if they engage with the task very superficially.

An example of a 'difficult' question on a comprehension passage where the expected response is open-ended is: Is the order in which the various sports are dealt with in the passage as a whole significant? Justify your answer. (3 marks).

An example of a 'difficult' question on a prescribed novel requiring a relational response and involving demanding marking strategies (expected response difficulty) is: George Orwell uses satire to explore human evil. In a well-constructed essay of 400–450 words (2–2½ pages), discuss to what extent this statement applies to *Animal Farm*. (Satire: the use of humour, irony, exaggeration or ridicule to expose and criticize people's stupidity or vices, particularly in the context of contemporary politics and other topical issues.)

The following are examples of probable invalid sources of expected response difficulty:

- Mark allocation is unclear or illogical. The weighting of marks is important in questions that comprise more than one component when the different components vary in degree of difficulty. Students may be awarded the same marks for answering an easy component(s) of the item as other students are awarded for answering the more difficult components.

- Mark scheme and questions are incongruent.

- Question asked is not the one that examiners want candidates to answer. Memorandum spells out expectation to a slightly different question, not the actual question.

- Impossible for candidate to work out from the question what the answer to the question is (answer is indeterminable).

- Wrong answer provided in memorandum.

- Alternative correct answers from those provided in the memorandum are also plausible.

- The question is 'open' but the memorandum has a closed response. Memorandum allows insufficient leeway for markers to interpret answers and give credit where due.

The category expected response difficulty thus also raises the importance of the quality of marking. Accuracy of marking is essential to the overall integrity of the degree of question difficulty (Sorto & Nádas, 2009).

## 3.3 Discussion

The 2011 Home Language report points to limitations in using pre-determined combinations of categories and concepts for making judgements about where the difficulty in a particular examination question might lie. The framework for thinking about item or question difficulty outlined in section 3.2 of this report does not provide evaluators with explicit links between the different categories and aspects of each category. Nevertheless, the assumption is that judgement of question difficulty is influenced by the interaction and overlap of a variety of elements in the different categories.

The difficulty framework assumes that evaluators can make these connections and assessments themselves. For example, evaluators need to decide whether a question that tests specific knowledge is actually more difficult that a multi-step question, because it requires candidates to explain a highly abstract theory, or very complex content. It is simply not possible to pre-determine, specify or show all possible permutations of relationships and overlaps between the different categories and concepts on the framework.

Leong (2006) stresses that his difficulty framework is not an explanatory theory; it does not provide an explanation why, for example, a question that tests specific knowledge can actually be more difficult than a multi-step, multiple-concept question. Furthermore, although the framework allows for the selection and combination of properties or qualities, the concepts in the framework and lists of examples of possible invalid sources of question difficulty are not intended to be definitive. In working with actual examination questions or

items, evaluators may find that they need to modify the framework and that concepts and elements in the framework need to be extended or improved.

Clearly a high level of expertise is needed to develop the kind of insight required to make connections at multiple levels across the different categories of difficulty in this framework. The tacit nature of this expert knowledge can limit its transfer to novice evaluators (Fisher-Hoch & Hughes, 1996). What the framework outlined in section 3.2 tries to provide, is a language for evaluators to articulate and share discussion about question difficulty. The difficulty framework provided could be used in ways that help build the capacity of novice and less experienced evaluators to exercise the necessary expert judgement by making them more aware of aspects they need to consider when judging question difficulty

## 3.4 New instrument for analysing Home Language examinations

The new instrument for Umalusi's comparative analysis of Home Language examinations can be found in **Appendix A** of this report.

First, what should be noted is that the instrument for analysing and comparing Home Language examinations draws a distinction between judging the level of cognitive demand of examination questions (Table A.1) and judging the difficulty level of examination questions (Table A.2).

Secondly, the addition of a forth level, 'very difficult', to the difficulty levels of examination questions in Table A.2 should be noted. Evaluators are now required to make judgments about whether each examination question makes low, medium, high or *very high* demands on the average Grade 12 examination candidate. They need to identify questions deemed to be 'very difficult' for the average Grade 12 student to answer, questions where skills and knowledge required to answer them allow for A-grade students (extremely high-achieving/ability students) to be discriminated from other high ability/proficiency students. The intention is to ensure that there are sufficient questions that discriminate well amongst higher ability candidates.

Thirdly, team members are required to use the difficulty framework (Table A.3) to help them identify and list the main sources of difficulty or 'easiness' (Content, Stimulus, Task and/or Expected Response) in each question, and to assist them in providing their reasoning for the level of difficulty selected for each examination question in Tables A4 and A5. Members first make their decisions on an individual basis (Table A.4). They record a consensus rating (Table A.5) after they have reached agreement through discussion with other members of their language teams.

Fourthly, in line with recommendations from Part 1 of the Home Language research report, in addition to judging the difficulty level of examination question, evaluators are also required to judge the difficulty of the stimulus material (reading selection or passage, visual text or source material including prescribed novels, poems, and dramas) that students are required to read or refer to in each of the Home Language examination papers. Teams of evaluators are asked to use ratings provided in Table A.6 to evaluate whether the stimulus or source material in each of the Home Language examination papers makes low, medium, or high 'reading' demands on the average Grade 12 examination candidate in Table A.7.

## 3.5  Conclusion

Umalusi's new framework is designed on the basis of the revised Bloom's and Barrett's Taxonomies and the conceptual framework outlined in this report. The new instrument provides examination evaluators as well as other key role players in the examination system such as examiners and moderators with a conceptual framework for thinking about item or question difficulty. The difficulty framework is intended to support all key role players in making complex decisions about what makes a particular question or item easy, moderately challenging, difficult or very difficult for Grade 12 examination candidates.

The difficulty framework could also be used as a means of increasing the awareness amongst teachers and students of the implicit demands residing in different categories of question difficulty, in particular in mark schemes. Ultimately, what the conceptual framework is intended to address, are concerns about whether the examinations in question are defensibly adequate instruments.

# Appendix A: Home Language examination paper analysis instrument

1. In this research, you are asked to analyse the 20… examination papers for your particular language. You will use the cognitive demand typology explained in Table A.1; and the levels of difficulty provided in Table A.2 and the framework for thinking about difficulty levels provided in Table A.3 to collect the data to complete Tables A.4 and 5 provided below; and the degrees of difficulty of source/stimulus material provided in Table A.6 to complete Table A.7 (below).

2. The findings from the investigation as described in point 1 above will serve as comparative data for the analysis of the final 20… NSC Home Language (HL) question papers. This analysis is planned to continue in a second workshop *immediately after* the final examination for your language has been written (October/November 20…), using the same Tables.

3. Initially you will be asked to use the tools and ratings provided to conduct an item-by-item analysis of examination papers independently of other team members in your language group. In order to establish negotiated decisions regarding the item/ question analysis, the particular team members will then have to work together in order to determine the nature of the cognitive demands made in the Home Language papers and the comparable degree of difficulty of the examinations. The comparison will be based on the analyses the teams have completed for the papers. Teams will also be asked to rate the difficulty levels of source or stimulus material in the Home Language examination papers.

4. Each team member has to assist the team leader to complete a report which reflects the analysis of the earlier examinations as well as your findings regarding the 20…. papers. This report will feed into the standardization process which Umalusi undertakes. Based on the findings which each team will submit, a comparison of the Home Language papers will be possible.

In **Round 1**, we ask you to independently rate each examination question and mark your ratings on Table A.4. We ask that, at this stage, you do not discuss your ratings with other team member in your group. The collection of data by doing an item-by-item analysis of the exam papers in Table A.4, serves as an individual preparatory exercise for each team member to work with the typology and levels of difficulty. Your independent judgments and comments will play a very important role in forming consensus decisions in Round 2.

Taxonomies of cognitive demand such as Bloom's Taxonomy suggest that cognitive processes exist in a cumulative hierarchy. Please first read through the cognitive demand typology in Table A.1 and familiarize yourself with the levels and explanations of the types of categorizations. The cognitive demand taxonomy combines aspects of the Revised Bloom's Taxonomy and Barrett's Taxonomy and covers processes involved in comprehending written texts and in generating or creating texts.

**Table A.1: Typology of cognitive demand**

| Level of cognitive demand | Type of cognitive demand | Explanation of categorization. Question which require students: | Examples |
|---|---|---|---|
| **Lower order processes** | **1. Recognize or recall** | To locate, identify and retrieve any kind of **explicitly stated** information, ideas, facts or details in reading material provided, or from memory of previously learned or read material (for example, names of places), and recognition of the relevance of the information, ideas, facts or details in relation to the question | The contextual questions on Shakespeare's drama Romeo and Juliet: *Complete the following sentence by filling in the missing words. Write down only the question number and the words. Juliet sends the Nurse to Friar Lawrence's cell to take Romeo a … and tell him to come to her that night and say …* The comprehension question: *Give two reasons why children become overweight. Refer to paragraph 3 (of the given passage).* |
| | **2. Apply or reorganize** | To use or **apply a basic procedure** (for example, a basic grammatical rule), **to replicate a model or version** (for example, a basic visual representation, a report, memo, invitation in a highly scaffolded way where students have to recreate rather than create), or to **reorganize explicitly stated information**, ideas, facts or details from reading material or from memory of previously learned or read material in a different way or form from what was presented (for example, to sort, classify, match, categorize, compare, contrast, summarise or paraphrase, or consolidate **explicitly stated** information, ideas, facts or details.) | *Rewrite the following sentence in the passive voice starting with the given word: The 18-year-old had developed an illness causing paralysis. Start with: An …* *Rewrite the following sentence so that it is grammatically correct. 'When wearing their apparently sprayed-on outfits, it gives them a false sense of being stylish.'* |

| Level of cognitive demand | Type of cognitive demand | Explanation of categorization. Question which require students: | Examples |
|---|---|---|---|
| **Medium order processes** | **3.Infer, interpret or analyse** | To engage in **more abstract (inferential) reasoning and interpretation**, and use conjecture, background knowledge, clues or **implicit information**, ideas, facts or details in reading material provided or from memory of previously learned or read material as a basis of forming hypotheses, interpreting, inferring or analysing details, relationships or ideas (for example, the significance of a theme, the motivation or nature of a character) which are **not explicitly stated** in reading or other source material | The contextual questions on Shakespeare's drama *Romeo and Juliet: Juliet sends the Nurse to Romeo. What does this show the audience about the relationship between Juliet and the Nurse?*<br><br>The question on an extract from the novel *Animal Farm: Refer to lines 12–13: 'the animals crept silently away.' What do the underlined words convey about the animals' feelings at this stage of the novel?* |
| **Higher order processes** | **4. Evaluate or appreciate** | To make critical **judgement** (for example, on qualities of accuracy, consistency, acceptability, desirability, worth or probability) using criteria provided by other sources or authorities, or students' own values, experiences, or background knowledge of the subject<br>To show emotional and aesthetic or literary sensitivity or a **reaction to the worth of psychological and artistic elements** of reading material (including literary techniques, language, forms, styles, and structuring). (For example, commenting on the effectiveness of a poetic device or image). | The question on a *Madam and Eve* cartoon: *The cartoonist does not show the mother-in-law in any of the frames. Do you think that this is an effective technique? Justify your response.*<br><br>The question on an unseen poem, *An Abandoned Bundle by M. O. Mtshali: Discuss how the poet employs* **diction** *and* **imagery** *to reveal his state of mind to readers.* |
| | **5. Synthesise or create** | To **integrate** ideas and information and relate parts of material, ideas, or information to one another and to an overall structure or purpose in a way that is **relational**.<br>To engage in **original creative thought** and design and put elements together to form a coherent whole and make a new or **unique** product **showing emotional, aesthetic or literary sensitivity** | *You are selling a second-hand item (e.g. a Walkman, a CD player, an item of clothing). Create an advertisement which will be placed on the notice board at school. Write an essay of between 250 and 300 words titled 'As I looked at that photograph…'* |

It is important that a range of cognitive demands is made on students in Grade 12 examinations, and that examination papers are not only characterized by questions that require reproduction of information and replication of basic procedures. Questions that assess higher order processes, such as analysis, are *usually* more difficult than questions that assess lower order processes, such as recognition and factual recall (simple memory questions); hence higher level cognitive demands should not be under-represented in papers.

However, other factors besides the type of cognitive demand affect the degree of difficulty of examination questions for Grade 12 students. We know this, because questions that align to a particular type of cognitive demand are not always as difficult as other questions that align to the same cognitive demand. For example:

- a recall question that asks students to recall an abstract theory, or complex content is often much more difficult to accomplish than one which asks candidates to recall a simple fact (i.e. differences in content difficulty).

- a question that requires recalling an ordered sequence of events and which entails writing a few sentences is generally easier than one which involves the same type of cognitive demand but entails writing a whole essay (i.e. differences in 'writing difficulty').

- literal comprehension of source material comprising a simple contemporary magazine article is generally easier than literal comprehension of a classical work such as one of Shakespeare's plays because of differences in the content, vocabulary, sentence and organizational structure, register, diction, literary techniques, abstractness of ideas and imagery, and background knowledge required (i.e. differences in 'reading difficulty').

We also know that a particular evaluation question may be more difficult for a typical Grade 12 student than for a university graduate. A question requiring inferential reasoning may be far more difficult for the average Grade 6 student than for the average Grade 12 student. Hence, in addition to identifying the type of cognitive demand made in each examination question, what you are also asked to do in this investigation is to use your knowledge of the subject, your experience of teaching it, and/or marking Grade 12 examination scripts to make judgments about whether each examination question makes low, medium, high or very high demands on the average Grade 12 examination candidate.

Table A.2 provides a brief explanation and ratings for these four **levels of difficulty**.

**Table A.2: Degree of difficulty of examination questions**

| This question is assessed as: | | | |
|---|---|---|---|
| **1** | **2** | **3** | **4** |
| **Easy** for the average Grade 12 student to answer | **Moderately** challenging for the average Grade 12 student to answer | **Difficult** for the average Grade 12 student to answer | **Very difficult** for the average Grade 12 student to answer. The skills and knowledge required to answer the question allows for A-grade students (extremely high-achieving/ability students) to be discriminated from other high ability/ proficiency students |

When you rate the degree of the difficulty of a particular question or item for the **average** Grade 12 candidate, your conception of the average Grade 12 student should be representative of the entire population of Grade 12 candidates for all schools in the country, in other words, of the overall Grade 12 student population. When you think of the average Grade 12 candidate you should not reference only students from one area or region in the country, or only a certain group of schools or classrooms. You should reference the Grade 12 student population in general.

To judge the degree of difficulty of each examination question, you need to consider both the demands that each question makes on the cognitive schema of an average Grade 12 Home Language examinee and the intrinsic difficulty of the question or task. To make this judgment, you need to identify where the difficulty or ease in a particular question resides.

**Table A.3** provides you with a framework for thinking about item or question difficulty. The framework gives guidance on where the difficulty in a question might reside. This guidance is intended to support you in making complex decisions about what makes a particular question or item easy, moderately challenging, difficult or very difficult for Grade 12 examination candidates.

The four-category framework *for thinking about question or item difficulty* in Table A.3 is adapted from Leong (2006) and comprises the following four general categories of difficulty:

- Content (subject/concept) difficulty;

- Stimulus (question) difficulty;

- Task (process) difficulty; and

- Expected response difficulty.

When you read the framework you will see that

- the **demands of the reading required** to answer a question forms an important element of 'Stimulus difficulty'.

- the **demands of the written** text or representations students are required to produce for their response form an important element of 'Task difficulty'.

- the **cognitive demand** of a question is another element of 'Task difficulty'.

The framework draws a crucial distinction between 'valid' and 'invalid' sources of question difficulty in each of the four general categories. Valid difficulty or 'easiness' are those levels that are intended by the examiner. Invalid sources of difficulty are defined as those features of question difficulty that were *not* intended by the examiner. **Invalid sources of question difficulty** prevent the question from assessing what the examiner intended, and **are those aspects of question difficulty that prevent candidates from demonstrating their true ability or competence**. The framework provides **examples** of possible invalid or unintended sources for each category. For example, grammatical errors in a question that could cause misunderstanding are invalid sources of question difficulty because the difficulty in answering the question lies in the faulty formulation of the question, rather than in the intrinsic difficulty of the question itself. Students may misunderstand the question and therefore not be able to demonstrate what they know.

Table A.3 provides the four-category conceptual framework adapted from Leong (2006).

**Table A.3: Framework for thinking about question difficulty**

| CONTENT/CONCEPT DIFFICULTY |
|---|
| **Content/concept difficulty** indexes the difficulty in the **subject matter, topic or conceptual knowledge** assessed or required. In this judgment of the item/question, difficulty exists in the **academic and conceptual demands** that questions make and/or the **grade level** boundaries of the various 'elements' of domain/subject knowledge (topics, facts, concepts, principles and procedures associated with the subject). |
| **For example** |
| Questions that assess 'advanced content', that is, subject knowledge that is considered to be in advance of the grade level curriculum, are *likely* to be difficult or very difficult for most candidates. |
| Questions that assess subject knowledge which forms part of the core curriculum for the grade are *likely* to be moderately difficult for most candidates. |
| Questions that assess 'basic content' or subject knowledge candidates would have learnt at lower grade levels, and which would be familiar to them are *unlikely* to pose too much of a challenge to most candidates. |
| Questions that require general everyday knowledge or knowledge 'real life' experiences are *often* easier than those that test more specialized school knowledge. |
| Questions involving only concrete objects or phenomena are *usually* easier than those that involve more **abstract constructs or ideas**. For example, questions which test students' understanding of theoretical or **de-contextualized issues or topics**, rather than their knowledge of specific examples or contextualised topics or issues tend to be more difficult. |
| Content difficulty may also be varied by changing **the number of knowledge elements assessed**. *Generally* the difficulty of a question increases with the number of knowledge elements assessed. |
| Questions that assess students on two or more knowledge elements are *usually* (but not always) more difficulty than those that assess a single knowledge element. |
| Assessing students on **a combination of knowledge elements that are seldom combined** *usually* increases the level of difficulty. |
| EXAMPLE QUESTION: CONTENT/CONCEPT DIFFICULTY |
| Examples of 'difficult' questions involving more abstract ideas are: |
| Comment on the suitability of the title of the poem *The Weeping of the Penny Whistle* by Peter Horn. |
| Explain how humour is created in the *Madam and Eve* cartoon (provided). |
| EXAMPLES OF INVALID OR UNINTENDED SOURCE OF CONTENT DIFFICULTY |
| • Testing obscure or unimportant concepts or facts that are hardly mentioned in the curriculum, or which are unimportant to the curriculum learning objectives. |
| • Testing advanced concepts that candidates are extremely unlikely to have had opportunities to learn. |

| STIMULUS DIFFICULTY |
|---|
| **Stimulus difficulty** refers to the difficulty of the linguistic **features of the <u>question</u>** (**linguistic** complexity) and the challenge that candidates face when they attempt to read and understand the words and phrases in the question AND when they attempt to read and understand the <u>**information or 'texts'** (**diagrams, tables and graphs, pictures, cartoons, passages, etc.**) <u>that accompany the question</u>.</u> |
| **For example** |

Questions that contain words and phrases that require only simple and straightforward comprehension are *usually* easier than those that require the candidate to understand **subject specific phraseology and terminology** (e.g. idiomatic or grammatical language not usually encountered in everyday language), or that require more technical comprehension and specialized command of words and language (e.g. everyday words involving different meanings within the context of the subject).

Questions that contain information that is tailored to an expected response, that is, that contain no irrelevant information, are *generally* easier than those than require candidates to select relevant and appropriate information or **unravel a large amount of information** for their response.

Questions that depend on reading and selecting content from a text (including a graph, picture, cartoon, etc.) can be more challenging than questions that do not depend on *actually reading the text* because they test reading comprehension skills as well as subject knowledge.

Questions that require candidates to **read a lot** *can* be more challenging than those that require limited reading. Although the degree of difficulty in examinations is *usually* revealed most clearly through the questions, text complexity or the degree of **challenge in written or graphic texts** that students are required to read and interpret in order to respond can increase the level of difficulty. The degree of difficulty *may* increase if texts set, and reading passages or other **source material** used are challenging for the grade level, and make **high reading demands** on students at the grade level. Predictors of textual difficulty include

- **semantic content** – for example, if vocabulary and words used are typically outside the reading vocabulary of Grade 12 students, 'texts' (passage, cartoon, diagram, table, etc.) are *usually* more difficult. If words/images are made accessible by using semantic, syntactic or graphophonic cues, 'texts' are *generally* easier.

- **syntactic or organizational structure** – for example, sentence structure and length. For example, if students are likely to be *familiar with the structure* of the 'text', for example, from reading newspapers or magazines, etc. 'texts' are *usually* easier than when the structure is unfamiliar.

- **literary techniques** – for example, abstractness of ideas and imagery and background knowledge required, for example, to make sense of allusions. For example, if the context is unfamiliar and candidates do not have access to the context which informs a text (passage, cartoon, diagram, table, etc.) they are expected to read, and which informs the question they are supposed to answer and the answer they are expected to write, then constructing a response is *likely* to be more difficult than when the context is familiar.

Another factor in stimulus difficulty is presentation and visual appearance. For example, type face and size, use of headings etc. can aid 'readability' (Mobely, 1987 in Fisher-Hoch & Hughes, 1996).

| EXAMPLE QUESTION: STIMULUS DIFFICULTY |
|---|
| An example of a 'difficult' question because of the demands involved in unravelling the question is: |
| Your school is preparing a pamphlet to be given to the new Grade 8 learners. You have been asked to write, for inclusion in this pamphlet, instructions on how to manage the challenges of being in high school. Write out these instructions. |
| **EXAMPLES OF INVALID OR UNINTENDED SOURCES OF STIMULUS DIFFICULTY** |
| • Meaning of words unclear or unknown |
| • Difficult or impossible to work out what the question is asking |
| • Questions which are ambiguous |
| • Grammatical errors in the question that could cause misunderstanding |
| • Inaccuracy or inconsistency of information or data given |
| • Insufficient information provided |
| • Unclear resource (badly drawn or printed diagram, inappropriate graph, unconventional table) |
| • Dense presentation (too many important points packed in a certain part of the stimulus). |

| TASK DIFFICULTY |
| --- |
| **Task difficulty** refers to the **difficulty that candidates confront when they try to formulate or produce an answer**. |
| **For example** |

In most questions, to generate a response, candidates have to work through the steps of a solution. *Generally*, questions that **require more steps in a solution** are more difficult than those that require fewer steps.

Task difficulty may also be mediated by the **amount of guidance present in the question**. Although question format is not necessarily a factor and difficult questions can have a short or simple format, questions that provide guided steps are *generally* easier than those that are more open ended and require candidates to form their own response strategy, work out the steps and maintain the strategy for answering the question by themselves. A high degree of prompting (a high degree of prompted recall, for example) *tends* to reduce difficulty level.

Questions that test specific knowledge are *usually* less difficult that multi-step, **multiple-concept questions**.

A question that requires the candidate to use a **high level** of appropriate **subject, scientific or specialised terminology** in their response tends to be more difficult than one which does not.

A question requiring candidates to **create a complex abstract (symbolic or graphic) representation** is *usually* more challenging than a question requiring candidates to create a concrete representation.

A question requiring writing a one word answer, a phrase, or a simple sentence is *often* easier to write than responses that **require more complex sentences, a paragraph or a full essay or composition**.

**Narrative writing** is *usually* easier than **writing discursively** (argumentatively or analytically). In subjects such as language(s) where one of the goals is that students learn to express themselves well (in English, etc.) and **writing skill** is part of what students are supposed to learn, some questions reflect expected response difficulty simply by 'creating the space' for A-grade candidates to demonstrate genuine insight, original thought or good argumentation, and to write succinctly and coherently about their knowledge.

In contrast, questions which require continuous prose or extended writing *may* also be easier to answer correctly or to get marks for than questions that require no writing at all or single letter answer (such as multiple choice), or a brief response of one or two words or short phrase(s), because **they test very specific knowledge**.

The **cognitive demand** or **thinking processes** required form an aspect of task difficulty. Some questions test thinking ability, and students' capacity to deal with ideas, etc. Questions that assess inferential comprehension or application of knowledge, or that require students to take ideas from one context and use it in another, for example, *tend* to be more difficult than questions that assess recognition or retrieval of basic information.

Questions requiring recall of knowledge are *usually* more difficult than questions that require simple recognition processes.

When the resources for answering the question are included in the examination paper, then the task is *usually* easier than when candidates have to use and select their own internal resources (for example, their own knowledge of the subject) to answer the question.

**EXAMPLE QUESTION: TASK DIFFICULTY**

An example of a 'difficult' comprehension question involving higher level reasoning:

*Disqualified (in South Africa) from local first-class cricket on the grounds of race, D'Oliviera (Basil D'Oliviera, a world-class talent who just happened to have the 'wrong' colour of skin) went to live in England in 1960, becoming one of the stars of the English team. When he was selected for a 1968 tour of South Africa, the apartheid government barred him – an act of folly that offended even the crustiest British conservatives, and turned South Africa into an international sporting pariah.*

What does 'When he was selected … crustiest British conservatives' suggest about the difference between British and South African conservatives at the time?

An example of a 'difficult' discursive essay involving opinion and reasoning is:

Should cellphones be used as a tool in education? Write an essay in which you clearly express your views on this topic.

**EXAMPLES OF INVALID OR UNINTENDED SOURCES OF TASK DIFFICULTY**

- Level of detail required in an answer is unclear

- Context is unrelated to or uncharacteristic of the task than candidates have to do

- Details of a context distract candidates from recalling or using the right bits of their knowledge

- Question is unanswerable

- Illogical order or sequence of parts of the questions

- Interference from a previous question

- Insufficient space (or time) allocated for responding

- Question predictability or task familiarity. If the same question regularly appears in examination papers or has been provided to schools as exemplars, students are likely to have had prior exposure, and practised and rehearsed answers in class (for example, when the same language set works are prescribed each year).

| EXPECTED RESPONSE DIFFICULTY |
| --- |

**Expected response difficulty** refers to difficulty imposed by examiners in a **mark scheme and memorandum**. This location of difficulty is more applicable to 'constructed' response questions, as opposed to 'selected' response questions (such as multiple choice, matching/true-false).

**For example**

When examiners expect few or no details in a response, the question is *generally* easier than one where the mark scheme implies that **a lot of details are expected**.

Another element is the **complexity in structure of an expected response**. When simple connections between ideas are expected in a response, the question is *generally* easier to answer than a question in which the significance of the relations between the parts and the whole is expected to be discussed in a response. In other words, a question in which an unstructured response is expected is *generally* easier than a question in which **a relational response** is expected. A response which involves combining or linking a number of complex ideas is *usually* more difficult than a response where there is no need to combine or link ideas.

A further aspect of expected response difficulty is the **clarity of the allocation of marks**. Questions are *generally* easier when the allocation of marks is straight-forward or logical (i.e. 3 marks for listing 3 points) than when the **mark allocation is indeterminate** (e.g. when candidates need all 3 points for one full mark or 20 marks for a discussion of a concept, without any indication of how much and what to write in a response). This aspect affects difficulty, because candidates who are unclear about the mark expectations in a response may not produce sufficient amount of answers in their response that will earn the marks that befit their ability.

Some questions are more difficult/easy to mark accurately than others. Questions that are **harder to mark and score objectively** are *generally* more difficult for candidates than questions that require simple marking or scoring strategies on the part of markers. For example, recognition and recall questions are *usually* easier to test and mark objectively because they usually require the use of matching and/or simple scanning strategies on the part of markers. More complex questions requiring analysis (breaking down a passage or material into its component parts), evaluation (making judgments, for example, about the worth of material or text, or about solutions to a problem), synthesis (bringing together parts or elements to form a whole), and creativity (presenting original thought) are *generally* harder to mark/score objectively. The best way to test for analysis, evaluation, synthesis and creativity is usually through essays. Such essays *generally* require the use of more cognitively demanding **marking** strategies such as interpreting and evaluating the logic of what the candidate has written.

Questions where **a wide range of alternative answers or response(s)** is possible *tend* to be more difficult. In contrast, questions may be so open-ended that students will get marks even if they engage with the task very superficially.

| EXAMPLE QUESTION: EXPECTED RESPONSE DIFFICULTY |
| --- |

An example of a 'difficult' question on a comprehension passage where the expected response is open-ended is: Is the order in which the various sports are dealt with in the passage as a whole significant? Justify your answer. (3 marks)

An example of a 'difficult' question on a prescribed novel requiring a relational response and cognitively demanding marking strategies is: George Orwell uses satire to explore human evil. In a well-constructed essay of 400–450 words (2–2½ pages), discuss to what extent this statement applies to *Animal Farm*. (Satire: the use of humour, irony, exaggeration or ridicule to expose and criticize people's stupidity or vices, particularly in the context of contemporary politics and other topical issues.)

| EXAMPLES OF INVALID OR UNINTENDED SOURCES OF EXPECTED RESPONSE DIFFICULTY |
|---|
| • Mark allocation is unclear or illogical. The weighting of marks is important in questions that comprise more than one component when components vary in degree of difficulty. Students may be able to get the same marks for answering an easy component(s) of the item as other students are awarded for answering the more difficult components |
| • Mark scheme and questions are incongruent |
| • Question asked is not the one that examiners want candidates to answer. Memorandum spells out expectation to a slightly different question, not the actual question |
| • Impossible for candidate to work out from the question what the answer to the question is (answer is indeterminable) |
| • Wrong answer provided in memorandum |
| • Alternative correct answers from those provided or spelt out in the memorandum are also plausible |
| • The question is 'open' but the memorandum has a closed response. The memorandum allows no leeway for markers to interpret answers and give credit where due. |

The above framework does not provide you with explicit links between the different categories, or show relationships between the different concepts in the framework. This is because it is impossible to set prescribed rules or pre-determined combinations of categories and concepts for making judgements about where the difficulty in a particular examination question might lie.

The idea behind the framework is to allow you to exercise your sense of judgement as a coherent whole. The complexity of your judgement task lies in your ability as an expert to recognize subtle interactions and identify links between different elements of a question's difficulty or 'easiness'. For example, a question that tests specific knowledge can actually be more difficult that a multi-step question, because it requires candidates to explain a highly abstract theory, or very complex content. The framework is also intended to assist you in justifying your decisions with regard to difficulty level and invalid difficulty ratings.

The concepts in the framework and the examples of invalid sources of question difficulty are not intended to be definitive. It may be that, in working with actual examination items you find you need to add other elements and examples to the framework. Please feel free to add your own ideas. Remember to keep a record of your ideas and your suggestions for additions or amendments to the framework and examples as you justify your ratings when you complete **Table A.4**.

While you are working on your examination item analysis, also give some thought to the best use of the examination analysis tool overall. Keep a note of your thoughts. There will be time to discuss your comments during the workshops. Your ideas will be included in the reporting and considered for inclusion in future examination analysis instruments.

Additional guidance for completing **Table A.4** is given directly below the table.

**Table A.4: Table for recording of individual analysis of NSC examination paper questions**

| Note: This table is to be completed by each team member individually | | | | | | |
|---|---|---|---|---|---|---|
| 20... ENG P1 | | | | | | |
| C1 | C2 | C3 | C4 | C5 | C6 | C7 |
| Item | Marks | Type of cognitive demand (1,2,3,4,OR 5) | Difficulty level (1, 2 3, 4 or ID) | Identify and list the main sources of difficulty (Content, Stimulus, Task and/or Expected Response) | Justify/ defend your difficulty level or ID rating | Content / skill/ topic |
| 1.1 | | | | | | |
| 1.2 | | | | | | |
| 1.3 | | | | | | |
| 1.4 | | | | | | |
| 1.5 | | | | | | |
| Etc | | | | | | |
| | | | | | | |
| | | | | | | |
| TOTAL | 70 | | | | | |

*(Please add as many rows as necessary)*

**Guidance to complete the individual analysis in table A.4**

1. Enter the *examination paper code* at the top left hand corner of Table A.4 (indicate in abbreviated form, the date, the subject, and paper number (1, 2, 3), for example, 20.. Eng P1.

2. Column 1 (C1): Place the *question number, or the subsection number* in the '**item**' column.

3. Column 2 (C2): Enter the mark for that item in the second column, marked '**marks**'.

4. Column 3 (C3): Enter a cross in the column representing the **type of cognitive demand** you have identified. Refer to Table A.1 above. The 5 different types of cognitive demand are represented by the following abbreviations:

    1. Recognize or recall

    2. Apply or reorganize

    3. Infer, interpret or analyse

    4. Evaluate or appreciate

    5. Synthesise or create.

5. Column 4 (C4): Rate the **degree of difficulty** of each item/question. Refer to Table A.2 above. Enter for instance a **1** in the column when you are of opinion that the question or item would be **easy** for the average Grade 12 student to answer; enter a **2** in the column for items that would be **moderately challenging** for the average Grade 12 student to

answer, enter **3** in the column for items or questions that would be **difficult** for the average Grade 12 student to answer, and enter a **4** if the skills and knowledge required to answer the question are **very difficult** and allow for A-grade students (extremely high-achieving/ability students) to be discriminated from other high ability/proficiency students. (This may even be simply by 'creating the space' for A-grade candidates to demonstrate genuine insight, and to write succinctly and coherently about their knowledge.) Enter **ID** if you have assessed a question or item as having an invalid source of question difficulty.

6. Column 5 (C5). Refer to the framework for thinking about question difficulty (Table A.3) and identify the characteristic(s) which you think make the question easy, moderately difficult, difficult or very difficult for students at this level (i.e. Content difficulty, Stimulus difficulty, Task difficulty and/or Expected Response difficulty). List the relevant categories in column 5.

7. Column 6 (C6): You must justify or defend your decision and explain what it is about each category of question difficulty you have identified and listed in C5 that led you to award the particular difficulty rating. Refer to the framework (Table A.3) for possible reasons. If you have assessed a question or item as having an **invalid source** (ID) of question difficulty, you must explain why this is the case. Refer to the framework for possible reasons and examples of invalid difficulty for each category on the framework.

8. Column 7 (C7): write the **content/ skill area or topic** which is linked to the question.

Once you have completed the individual analysis of the 20…–20… examination papers (Table A.5), in **Round 2**, you are given the opportunity to discuss and revise your rating for each question in light of the discussion with the other members of you language team.

### Table A.5: Table for recording team analysis of NSC exam paper questions

| 20… ENG P1 | | | | | | |
|---|---|---|---|---|---|---|
| C1 | C2 | C3 | C4 | C5 | C6 | C7 |
| Item | Marks | Type of cognitive demand (1,2,3,4,OR 5) | Difficulty level (1, 2 3, 4 or ID) | List the main sources of difficulty (Content, Stimulus, Task and/or Expected Response) | Justify/ defend your difficulty level or ID rating | Content / skill/ topic |
| 1.1 | | | | | | |
| 1.2 | | | | | | |
| 1.3 | | | | | | |
| 1.4 | | | | | | |
| 1.5 | | | | | | |
| Etc | | | | | | |
| | | | | | | |
| | | | | | | |
| TOTAL | 70 | | | | | |

*(Please add as many rows as necessary)*

**Guidance to complete the team analysis in table A. 5**

Begin by having each member of the team tell the others what their individual decision for cognitive demand and degree of difficulty was for each question. Then discuss differences in the ratings and your various reasons for the ratings. When a convergence of decisions is reached in the group, record your consensus ratings and your justification for difficulty ratings and for invalid sources of question difficulty.

Remember that your individual knowledge and experience, and independent judgments and explanations are just as important as the views of influential members of the team in the consensual decision-making process. Your group decision-making in awarding ratings should not be easily swayed merely by the status or personalities of different evaluators.

Table A.5 requires you to identify, discuss and decide as a group on the particular information and to report on the following per question (or per subsection of the question, whichever is the smaller of the units). With reference to Table A.5:

1. Enter the **examination paper code** at the top left hand corner of Table A.5 (above) (indicate in abbreviated form, the date, the subject, and paper number (1, 2, 3), for example, 20.. Eng P1.

2. Column 1 (C1): Write the *question number, or the subsection number* in the 'item' column. If, in your opinion, one or more subsections of a question could stand independently, please treat each as a separate item in the table.

3. Column 2 (C2): Enter the mark for that item in the column, marked '**marks**'. (You should be able to cut and paste Columns 1 and 2 from your previous analysis.).

4. Column 3 (C3): With reference to your individual analysis (Table A.4) regarding the type of cognitive demand, discuss and agree in your groups on the **type of cognitive demand** required in each question by entering one of the following abbreviations on Table A.5:

    1. Recognize or recall

    2. Apply or reorganize

    3. Infer, interpret or analyse

    4. Evaluate or appreciate

    5. Synthesise or create.

5. Column 4 (C4): With reference to your individual analysis (Table A.4) regarding the **degree of difficulty**, discuss and agree in your groups on the degree of difficulty of each item/question. Use a scale of 1 – 4, where **1 (Easy)** represents simple and basic items; **2 (Moderate)** is for items of average difficulty, **3 (Difficult)** for difficult items involving more sophisticated linguistic competence, and **4 (Very difficult)** for items which allow for A-grade students to be discriminated from other high ability/proficiency students. Enter **ID** if you assess a question or item as having an **invalid source** of question difficulty.

6. Column 5 (C5): Your team must decide on the characteristic(s) from the framework which you think make the question easy, moderately difficult, difficult or very difficult for students at this level (i.e. Content difficulty, Stimulus difficulty, Task difficulty and/or Expected Response difficulty). List the relevant categories in column 5.

7. Column 6 (C6): Your team must justify or defend your decisions and explain what it is about each category of question difficulty you have identified and listed in C5 that led you to award the particular difficulty rating. For example, you must explain why a **Very difficult** question allows for A-grade students (extremely high-achieving students) to be discriminated from other high ability/proficiency students. Refer to the framework

for *possible* reasons. If the team has assessed a question or item as having an **invalid source** of question difficulty (ID), the team must explain why this is the case. Refer to the framework (Table A.3) for possible reasons and examples of invalid difficulty for each category on the framework.

8. Column 6 (C6): Agree in the group on the **content/ skill area/ topic** which is/ are linked to the question.

In addition to making judgments about whether each examination question makes low, medium, high or very high demands on the average Grade 12 examination candidate, what teams are also asked to do, is to make judgments about whether the source or stimulus material in each of the HL examination papers makes low, medium, or high reading demands on the average Grade 12 examination candidate.

For each new text (reading selection, visual text or source material including prescribed novels, poems, and dramas) that students are required to read *or refer to* in each of the Home Language examination papers, your teams need to discuss and rate the demands of the 'text' using the ratings provided **Table A.6** below. **Table A.6** provides a brief explanation and ratings for **three difficulty levels**.

## Table A.6: Degrees of difficulty of source/stimulus material

| **Content, vocabulary, sentence and organizational structure, register, diction, literary techniques, abstractness of ideas and imagery, and/or background knowledge required** for the reading selection, visual text or source material (including prescribed novels, poems, and drama) is assessed as: | | |
|---|---|---|
| **1** | **2** | **3** |
| Simple/easy for the average Grade 12 student to read and understand | Moderately challenging but accessible for the average Grade 12 student to read and understand | Difficult/complex for the average Grade 12 student to read and understand |

Guidance for completing **Table A.7** is given directly below the table.

## Table A.7: Table for recording team analysis of source or stimulus material in the NSC examination papers

| **20… ENG P1** | | |
|---|---|---|
| **C1** | **C2** | **C3** |
| **Item** | **Difficulty level (1,2,3)** | **Justify your decision by identifying the main source(s) of difficulty for Grade 12 candidates (content; vocabulary; sentence and organizational structure; register; diction; literary techniques; abstractness of ideas and imagery; and/or background knowledge required)** |
| | | |
| | | |
| | | |
| | | |

*(Please add as many rows as necessary)*

**Guidance to complete the team analysis of source/stimulus material in table A.7**

Table A.7 requires you to identify, discuss and decide *as a group* on the particular information and to report on the following per source/stimulus material.

1. Enter the ***examination paper code*** at the top left hand corner of Table A.7 (above) (indicate in abbreviated form, the date, the subject, and paper number (1, 2, 3), for example, 20.. Eng P1.

2. Column 1 (C1): Write the *question number,* or the *subsection number* of the relevant source/stimulus material in this column.

3. Column 2 (C2): With reference to your analysis regarding the degree of **difficulty** of each 'text'. Use the scale of 1 – 3, where **1** represents **Easy**; **2** represents **Moderately** challenging; and **3** represents **Difficult**.

4. Column 3 (C3): Your team must justify or defend your decisions and explain what it is about each text that led you to award the particular difficulty rating.

# References

Ahmed, A., and Pollitt, A. (1999). Curriculum demands and question difficulty. Paper presented at IAEA Conference, Slovenia.

Ahmed, A, Pollitt, A., Crisp, V., and Sweiry, E. (2003). *Writing examinations questions*. A course created by the Research & Evaluation Division, University of Cambridge Local Examinations Syndicate.

Allington, R. (1991). Effective literacy instruction for at-rIsk children. In *Better schooling for the children of poverty: Alternatives to conventional wisdom, volume II*. Eds. M. S. Knapp and P. M. Shields. Berkley: McCrutchen.

Anderson, L. and Sosniak. L. (1994). *Bloom's taxonomy: A forty-year retrospective*. Chicago, Illinois: The University of Chicago Press.

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., and Wittrock, M. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. (Abridged edition). New York: Addison Wesley Longman.

Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508-600). Washington, DC: American Council on Education.

Barrett, T. C. (1968). What is reading? Some current conceptions. In *Innovation of the National Society for the Study of Education*. Ed. H.M. Robinson. Chicago: University of Chicago Press.

Barrett, T. C. (1976). Taxonomy of reading comprehension. In *Teaching reading in the middle grades*. Eds. R. Smith and T. C. Barrett. Reading. MA.: Addison-Wesley.

Bernstein, B. (1999). Vertical and horizontal discourse: an essay. *British Journal of Sociology of Education*, 20 (2):157-173.

Biggs, J.B. (1999). *Teaching for quality learning in university*. Buckingham: Society for Research in Higher Education and Open University Press.

Biggs, J.B., and K.F. Collis. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Bloom B. S. (Ed.) Englehart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay.

Booker, M. J. (2007). A roof without walls: Benjamin Bloom's taxonomy and the misdirection of American education. *Academic Questions*, 20:347-355. Springer Science and Business Media.

Buckendahl, C. W., Smith R. W., Impara, J. C. and Plake, B. S. (2002). A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational Measurement*, 39 (3): 253-263.

Champeau De Lopez, C. Marchi, G. B and Arreaza-Coyle, M. E. (1997). A taxonomy: Evaluation reading comprehension in EFL. *Forum*, 35 (2): 30-38 http://eca.state.gov/forum/vols/vol35/no2/p30.htm (accessed 7 August 2011).

Cresswell, M. J. (1997). Examining judgements: Theory and practice of awarding public examination grades. Unpublished doctoral dissertation, Institute of Education, University of London, UK.

Crisp, V. (2010). Judging the grade: exploring the judgement processes involved in examination grading decisions. *Evaluation & Research in Education*, 23 (2): 19-35.

Department of Education. (2002). *Revised National Curriculum Statement Grades R to 9 (Schools) Policy*. Pretoria: Department of Education.

Department of Education (2008a). *National Curriculum Statement Grades 10-12 (General). Subject Assessment Guidelines Languages: English, Home Language, First Additional Language, Second Additional Language, January 2008*. Pretoria: Department of Education.

Department of Education (2008b). *National Curriculum Statement Grades 10-12 (General). Learning Programme Guidelines Languages: English, Home Language, First Additional Language, Second Additional Language, January 2008*. Pretoria: Department of Education.

Department of Education (2009a). *Examination Guidelines, Grade 12, Literature: Paper 2: English, Home Language and First Additional Language*, and Paper 1: Section D: Second Additional Language. Pretoria: Department of Education.

Department of Education (2009b). *Guidelines for the setting of Grade 12 examinations: Paper 1: English, Home Language, First Additional Language, Second Additional Language*. Pretoria: Department of Education.

Department of Education (2009c). Guidelines for the setting of Grade 12 examinations: Paper 3: English, Home Language, First Additional Language, Second Additional Language. Pretoria: Department of Education.

DuBay, W.H. (2004). The principles of readability. Costa Mesa, CA: Impact Information. Downloaded from http://www.nald.ca/fulltext/readab/readab.pdf (accessed 11 August 2011).

Edwards, J and Dall-Alba, G. (1981). Development of a scale of cognitive demand for analysis of printed secondary science materials. *Research in Science Education*, 11, 158-170.

Fisher-Hoch, H., and Hughes, S. (1996). What makes mathematics exam questions difficult? Paper presented at the British Educational Research Association Annual Conference, Lancaster University.

Fisher-Hoch, H., Hughes, S., and Bramley, T. (1997). What makes GCSE examination questions difficult? Outcomes of manipulating difficulty of GCSE questions. Paper presented at the British Educational Research Annual Conference, University of York.

Gray, W. S. (1960). The major aspects of reading. In *Sequential development of reading abilities*. Ed. H. M. Robinson. Supplementary Educational Monographs No. 90. Chicago: University of Chicago Press.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18 (4): 277-296.

Halliday M.A.K. (1973). *Explorations in the functions of language*. London: Edward Arnold.

Halliday M.A.K. (1978). *Language as social semiotic*. London: Edward Arnold.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education*, 62 (2):143-158.

Heins, S. and Skaggs, G. (2010). Conceptualizing the classroom of target students: A qualitative investigation of panelists' experiences during standard setting. *Educational measurement: Issues and practice*, 29 (2): 36-44.

Herber, H. (1978). *Teaching reading in content areas*. 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall.

Hess, K. (2005). Exploring cognitive demand in instruction and assessment. *National Center for the Improvement of Educational Assessment*. http://www.604and609.org/uploads/4/6/9/6/4696562/dok_applyingwebb_kh08.pdf (accessed 17 August 2011).

Hess, K., Carlock, B., Jones, B. and Walkup J.R. (2009). What exactly do "fewer, clearer, and higher standards" really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments.

http://www.nciea.org/publications/cognitiverigorpaper_KH11.pdf (accessed August 2011). Status: 'This article is currently pending publication. Permission to reproduce is given when authorship is fully cited.'

Hosseini, J. (1993). Application of Bloom's taxonomy and Piaget's model of cognitive processes to teaching of management information systems concepts, *Journal of Education Systems*, 5 (3): 1-12.

Hugo, W, Bertram, C, Green, W and Naidoo, D. (2008). Bernstein, Bloom and the analysis of pedagogy in South African Schools. *Journal of Education*, 43, 31-56.

Impara, J. C. and Plake, B.S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34, 353-366.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice*, 41 (4): 212-219.

Krathwohl, D.R., Bloom, B. and Masia. B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain*. London: Longman. pp. 143-156.

Leong, S.C. (2006). On varying the difficulty of test items. Paper presented at the 32 Annual Conference of the International Association for Educational Assessment, Singapore. http://www.iaea.info/documents/paper_1162a1d9f3.pdf (accessed August 2011).

Moll, L. C. (1991). Social and instructional Issues in literacy instruction for "disadvantaged" students. In *Better schooling for the children of poverty: Alternatives to conventional wisdom, volume II*. Eds. M. S. Knapp and P. M. Shields. Berkley: McCrutchen.

Muller, J, (2001). *Reclaiming knowledge: social theory, curriculum and education policy*. London: Routledge.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., Trong, K.L., and Sainsbury, M. (2009). PIRLS 2011 Assessment Framework. TIMMS & PIRLS International Study Center. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Newman, J.H. (1993). The critical thinking movement in historical perspective. In *Critical thinking: What every person needs to survive in a rapidly changing world*. Ed. R. Paul. Rohnert Park. California: Sonoma State University Press.

Nicholls J.G. and Muller, A. T. (1983). The differentiation of the concepts of difficulty and ability. *Child Development*, 54 (4): 951-959.

Osterlind, S. J. (1990). Toward a uniform definition of a test item. *Educational Research Quarterly*, 14 (4): 2-5.

Paul, R. (1993). *Critical thinking: What every person needs to survive in a rapidly changing world* (3rd ed.). Rohnert Park, California: Sonoma State University Press.

Pearson P. D., and Johnson, D.D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Phillips, D. C., and Kelly, M. E. (1975). Hierarchical theories of development in education and psychology. *Harvard Educational Review*, 45, 351-375.

Pollitt, A., Huchinson, C., Entwistle, N., & De Luca, C. (1985). *What makes exam questions difficult?* Edinburgh, UK: Scottish Academic Press.

Porter A. and J. Smithson. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In *From Capitol to the Classroom: Standards-Based Reform in the States*. Ed. S. Fuhrman. Chicago: National Society for the Study of Education, University of Chicago.

Reeves, C. ( 2005). The Effect of 'Opportunity-to Learn' and Classroom Pedagogy on Mathematics Achievement in Schools Serving Low Socio-Economic Status Communities in the Cape Peninsula. Unpublished PhD dissertation. University of Cape Town.

Reeves, C. and Muller, J. (2005). Picking up the pace: variation in the structure and organization of learning school mathematics. *Journal of Education*, 37, 1-28.

Sax, G., Ellenberg, E. G., and Klockars, A. J. (1972). Achievement as a function of test item complexity and difficulty. *Journal of Experimental Education*, 40 (4): 90-93.

Seddon, G. M. (1978). The properties of Bloom's taxonomy of education objectives for the cognitive domain. *Review of Educational Research*, 48 (2): 303-323.

Skagg, G. and Hein, S. (2011). Reducing the cognitive complexity associated with standard setting: A comparison of the single-passage bookmark and yes/no methods. *Educational and Psychological Measurement*, 71 (3): 571-592.

Soden, R. and Pithers, R. (2001). Knowledge matters in vocational problem-solving: a cognitive view, *Journal of Vocational Education and Training* 53 (2): 205-22.

Stanovich, K. (1986). Mathew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21 (4): 360-407.

Suto, W. M. I. and Nádas, R. (2009). Why are some GCSE examination questions harder to mark accurately than others? Using Kelly's Repertory Grid technique to identify relevant question features. *Research Papers in Education*, 24 (3): 335-377.

Taylor, N. and Vinjevold, P. (Eds.) (1999). *Getting learning right: Report of the President's Education Initiative*. Johannesburg: Joint Education Trust.

Todd, T. Tindal, G. and Mccleery, J. (2006). Using concepts to frame history content. The Journal of Experimental Education, 74 (4): 331-49.

Umalusi. (2004). *Investigation into the standard of the Senior Certificate examination. A Report on Research Conducted by Umalusi*. S. Ndaba. Pretoria: Umalusi.

Umalusi. (2007). *Cognitive challenge. A report on Umalusi's research on judging standards of intended and examined curricula*. Pretoria: Umalusi.

Umalusi. (2009a). *2008 Maintaining Standards Report (English 1st additional Language, Geography, Life Science, Mathematics, Mathematical Literacy and Physical Science) From NATED 550 to the new National Curriculum, Part 3: Exam Paper Analysis*. H. Bolton. Pretoria: Umalusi.

Umalusi. (2009b). *2008 Maintaining Standards Report (English 1st additional Language, Geography, Life Science, Mathematics, Mathematical Literacy and Physical Science) From NATED 550 to the new National Curriculum, Part 2: Curriculum Evaluation*. H. Bolton. Pretoria: Umalusi.

Umalusi. (2009c). *2008 Maintaining Standards Report (English 1st additional Language, Geography, Life Science, Mathematics, Mathematical Literacy and Physical Science) From NATED 550 to the new National Curriculum, Part 1: Overview*. H. Bolton. Pretoria: Umalusi.

Umalusi. (2010). *2009 Maintaining Standards Report (Accounting, Business Studies, Economics, History,) Overview*. S. Grussendorff. Pretoria: Umalusi.

Umalusi. (2011). *Comparative Analysis of the National Senior Certificate Home Language*, 2008-2010 Afrikaans; English; isiNdebele; isiXhosa; isiZulu; and Siswati. B. Lepota and N.L. Thwala. Pretoria: Umalusi.

Vogler, K. (2005). Improve your verbal questioning. *The Clearing House, 79* (2): 98-102.

Wineburg, S. and Schneider, J. (2010). Was Bloom's Taxonomy pointed in the wrong direction? *Phi Delta Kappan*, 91 (4): 56-61.

SETTING STANDARDS IN EDUCATION AND TRAINING

10 YEARS

UMALUSI

Council for Quality Assurance in
General and Further Education and Training