

Avasha Rambiritch

University of Pretoria

Towards transparency and accountability: The story of the Test of Academic Literacy for Postgraduate Students (TALPS)

Abstract

Unfair tests, unfair testing methods and the use of tests to restrict and deny access have ensured a negative attitude to tests. The move in the recent past (Shohamy 2001, 2008; Fulcher & Davidson 2007; McNamara & Roever 2006) has been to promote the design and development of fair tests, by test developers who are willing to be accountable for their designs. A first step in this process is to ensure that every step of the design process is documented, and that this information is available to the public. Making this information available means that test takers are now equipped with information about the test and can now ask questions about the test. Test developers become real, not just experts 'hiding behind their designs'. Importantly, this kind of transparency ensures a channel of communication, not just between test developers and other experts in the field, but also between test developers and test takers. Applied linguists should strive to ensure

that the tests they design and use are fair, socially acceptable and have positive effects. This paper will illustrate that these concerns become important when one works within a framework that challenges test developers to consider questions related to every aspect of the test. In employing a framework that incorporates a concern for the empirical analyses of a test, as well as a concern for the social dimensions of language testing, one is compelled to ensure transparency and accountability in the testing process, as well as giving a voice to those often ignored, but most affected by the use of the test scores: the test-takers. In telling the story of the design and development of TALPS, this article is the first step in ensuring the transparency and accountability of the test developers of TALPS.

Keywords: academic literacy, construct, specifications, task types, sub-tests, descriptive statistics, transparency, accountability

1 Language and learning in South African tertiary institutions

Language remains a contentious issue in South Africa. The trauma of Bantu education still reverberates through the country. The effects will, no doubt, be felt for many years to come. The democratic attempt to right the wrongs of the past has made great strides in many areas, but has also created new challenges for which solutions need to be found:

In sum, the legacy of the past was a fractured system and a set of HEIs [Higher Education Institutions] bearing the scars of their origins. As South Africa entered a process of social, economic and political reconstruction in 1994, it was clear that mere reform of certain aspects of higher education would not suffice to meet the challenges of a democratic country aiming to take its place in the world. Rather, a comprehensive transformation of higher education was required, marking a fundamental departure from the socio-political foundations of the previous regime (CHE, 2004: 230).

One of these challenges remains the issue of language and learning. For many students who have been taught in their mother tongue, this is their first experience at being taught in English. Tertiary institutions, especially those considered previously advantaged, today need contingency measures to deal with this situation. Not accepting these students because of poor language proficiency would have simply been a repetition of the past. The trend has been to set up specific programmes to assist these students. Different institutions have, however, taken different routes. Some have set up academic support programmes, department and units, while others have offered degrees and diplomas on an extended programme system, where the programme is extended by a year to ensure that the relevant academic support is provided.

At the University of Pretoria, as at other universities in the country, poor pass rates and low student success are issues of concern. The university also attracts students from other parts of Africa and the world, lending even more diversity to an already diverse environment. Mdepa and Tshiwula (2012:27) note that in 2008 more than 9500 students from non-SADC African countries studied in South Africa. Very often these students do not have English as a first language. To ensure student success measures had to be put in place to assist these students. Fraser's (2008) view that "in order to overcome unjust social conditions, the institutionalised obstacles that are preventing participatory parity would need to be dismantled" is particularly relevant here. In this case an 'obstacle' to student success and participation could be the issue of under-preparedness in the language of instruction. At the University of Pretoria, measures were therefore put in place to assist students, particularly at postgraduate level.

For the last seven years the Academic Writing for Postgraduate Students (EOT 300) course has focused on helping to develop the academic writing needs of postgraduate students. There has, over this period, been an increasing demand for the course, as supervisors recognised the poor academic literacy levels of their students. This has been the focus of a study conducted by Butler (2007). The study focused on the design of a course for academic writing at tertiary level. He states that the "immediate context

of this study derives from the concern that a number of academic departments from a variety of disciplines at the University of Pretoria, have expressed about the academic writing ability (and general language proficiency) of their postgraduate students” (2007: 10). Butler (2007: 10) explains that these students are unfamiliar with academic writing conventions, are often unable to express themselves clearly in English, and have not “yet fully acquired the academic discourse needed in order to cope independently with the literacy demands of postgraduate study”. Information elicited from a questionnaire administered to students and supervisors and from personal interviews with supervisors as part of Butler’s study, confirmed that there are serious academic literacy problems experienced by these students, and that as a result of these problems students do not complete their studies in the required time. What became clear from the data derived from these questionnaires and the information from the interviews was the need for a “reliable literacy assessment instrument” (Butler, 2007: 181) that would “provide one with accurate information on students’ academic literacy levels” (2007: 181). The need for an academic literacy test for postgraduate students had thus been identified and work on the test began. This study is focused specifically on telling the story of the test i.e. on the design and development of the TALPS and in so doing, takes the first step towards ensuring transparency and accountability in the design and development of such tests.

2 Transparency and accountability in language testing

The concept of transparency has, in the recent past, become the watchword in government and politics, in the corporate world, in the media and even in the humanities and social sciences. Naurin (2007) states that transparency literally means that it is possible to look into something, to see what is going on:

A transparent organisation, political system, juridical process or market is one where it is possible for people outside to acquire the information they need to form opinions about actions and processes within these institutions (Naurin, 2007: 2).

True transparency means that information is easily available to those who need it and that, importantly also, this availability of information allows an open dialogue between those within and those outside of the organisation. The urgency of the need for transparency across all fields has led to the creation of a global body called Transparency International, an organisation that “brings together relevant players from government, civil society, business and the media to promote transparency in elections, in public administrations, in procurement and in business” (*Transparency International*, 2010). Similarly, the importance of transparency has been highlighted through the formation of the Global Transparency Initiative, which is a society “committed to the idea of greater openness at the international financial institutions” (*Global Transparency Initiative*, 2006). The concept of transparency in the field of language testing has not been explored in great detail. While experts in the field of language testing have stressed the need for an open dialogue between test developers and test takers, for test takers to be able to ask questions about the tests and for test developers to take responsibility for their

designs, this has not always been done. The first step towards achieving this begins with the concept of transparency because, realistically, one cannot ask questions about a process one knows nothing about, nor can one ask questions if one does not know who to direct these questions to. Weideman defines the transparency of a test as the “availability of information about its content and workings” (Weideman, 2006: 82).

The term ‘accountability’, like the term ‘transparency’, features prominently in the literature of many disciplines: commerce, law, education, public management and human resources (see Norton, 1997; Beu & Buckley, 2004), to name just a few. Explained simply, accountability has to do with taking responsibility for your actions. Accountability, however, does not stop there but requires, in addition to accounting for one’s actions, that one be willing to face the consequences of these actions. According to Sinclair (1995: 220), accountability entails a relationship in which people are required to explain and take responsibility for their actions. Bovens (2005: 7) explains that accountability should be defined as a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct to the forum, which then becomes a platform that can pose questions and pass judgment, and even sanction the actor. The same kind of relationship is echoed in other authorities. According to Frink and Klimoski (2004: 2), for example, definitions of accountability tend to

revolve around two specific themes. One theme concerns the context, that is, who and what is involved in a given situation, and the second theme involves the notion of an evaluation and feedback activity in some form (Frink & Klimonski, 2004: 3).

Explained simply: “Accountability involves an actor or agent in a social context who potentially is subject to observation and evaluation by some audience(s), including one’s self” (2004: 3). There are also

standards, or expectations against which the agent’s behaviour are compared, and the belief on the part of the agent of some likelihood that he or she may need to answer for, justify, or defend the decisions or behaviours. In addition, it is important that there are outcomes for the agent (i.e., sanctions, rewards, or punishments that can be explicit or implicit, and also objective or subjective) (Frink & Klimonski, 2004: 4).

In explaining his use of the term accountability, Weideman (2006: 72) turns to the definition provided by Schuurman (2005). Schuurman’s definition, too, stresses the need for one to be aware of his/her actions and to “give account of the same to the public” (Schuurman 2005: 42).

As applied linguists and test developers, these are some of the very issues that we should be concerned with i.e. that information about our tests are available to those most affected by the use of these test scores and that we be willing to take responsibility for our designs. These issues become relevant when one works within a framework that incorporates a concern for the empirical analyses of a test, as well as a concern for the

social dimension of language testing. In proposing such a framework, Weideman (2009) calls for a responsible agenda for applied linguistics, arguing that applied linguistic work should be backed by some foundational framework to ensure that the notions of responsibility, integrity, accessibility and fairness can be articulated in a theoretically coherent and systematic way. The framework he refers to is based on a “representation of the relationship among a select number of fundamental concepts in language testing” (Weideman, 2009: 241). The two main functions are the technical mode and the analytical dimension. These do not function in isolation. The relation between the two is ‘reciprocal’ (Weideman, 2009: 244). The technical mode interacts not only with the analytical mode, however, but is also connected with all other modes. Weideman points out that the technical unity of multiple sources of evidence, the reliability of a test, its validity and its rational justification are foundational or constitutive applied linguistic concepts (2009: 247). These may also be designated necessary requirements for tests (Weideman, 2009: 247). Important is the fact that “each of these ‘necessary’ or foundational concepts yields a (technically stamped) criterion or condition for the responsible use or implementation of the technical instrument” (Weideman, 2009: 247). This, according to Weideman, is why we say that tests should be reliable, valid and built on a theoretical base that is defensible in terms of a unity within a multiplicity of sources of evidence (Weideman, 2009: 247).

This technical dimension of the applied linguistic design also links with the lingual, social, economic, aesthetic, juridical and ethical aspects. According to Weideman, the links between the technical, qualifying function of the test design and other aspects yield the ideas of technical articulation, test implementation or use, technical utility, the alignment the test has with learning and teaching language, its public defensibility or accountability, and its fairness or care for those taking the test (Weideman, 2009: 247). In employing a framework that incorporates a concern for the empirical analyses of a test, as well as a concern for the social dimension of language testing one is compelled to ensure transparency and accountability in the testing process. In telling the story of the design and development of TALPS, this article is the first step in ensuring the transparency and accountability of the test developers of TALPS.

3 Telling the story of a test

3.1 Deciding on a construct

A first step for the developers was to find an appropriate construct on which to base the test and the intervention. Bachman and Palmer (1996:21) define a construct as the “specific definition of an ability that provides the basis for a given test or test task and for interpreting scores derived from this task.” The developers chose to base TALPS on the same construct as the Test of Academic Literacy Levels (TALL) (See Van Dyk & Weideman 2004). The TALL was in many ways a sounding board for TALPS – the success of TALL was one of the most important factors that motivated the development of TALPS. Both these tests are designed to test the same thing: the academic literacy

of students, the only difference being that one is directed at first year students while the other is intended for postgraduate students. The proposed blueprint for the test of academic literacy for the University of Pretoria, requires that students should be able to:

- understand a range of academic vocabulary in context;
- interpret and use metaphor and idiom, and perceive connotation, word play and ambiguity;
- understand relations between different parts of a text, be aware of the logical development of (an academic) text, via introductions to conclusions, and know how to use language that serves to make the different parts of a text hang together;
- interpret different kinds of text type (genre), and show sensitivity for the meaning that they convey, and the audience that they are aimed at;
- interpret, use and produce information presented in graphic or visual format;
- make distinctions between essential and non-essential information, fact and opinion, propositions and arguments; distinguish between cause and effect, classify, categorise and handle data that make comparisons;
- see sequence and order, do simple numerical estimations and computations that are relevant to academic information, that allow comparisons to be made, and can be applied for the purposes of an argument;
- know what counts as evidence for an argument, extrapolate from information by making inferences, and apply the information or its implications to other cases than the one at hand;
- understand the communicative function of various ways of expression in academic language (such as defining, providing examples, arguing); and
- make meaning (e.g. of an academic text) beyond the level of the sentence (Weideman 2003:61).

What is important here is the fact that the construct decided on “constitutes a definition of academic literacy” (Weideman 2003:61). The construct is a blueprint of what students should be able to do at tertiary level, and this is what the test tests. Weideman explains that in the process of developing the construct this view of academic literacy had been discussed at conferences and seminars and with “trans-disciplinary panels of academics” (2003:61). The responses have confirmed that “the elements identified above indeed constitute a number of essential components of what academic literacy entails” (2003:61). A further confirmation of the positive reception of the construct, according to Weideman,

came in the form of offers from other institutions to become partners in the development of the test or to use the test on their students (2003:61). The TALL also proved to be a highly valid and reliable test. In light of this the designers of TALPS were more than justified in using a blueprint that had already proved successful.

3.2 Specifications

The next step for the developers of TALPS was to align the construct of the test with specifications. Davies *et al.* (1999:207) define test specifications as a document which sets out what a test is designed to measure and how this will be tested. Davidson and Lynch (2002:4) state that “the chief tool of language test development is a test specification, which is a generative blueprint from which test items or tasks can be produced”. They state also that a well-written test specification can generate many equivalent test tasks. The discussion of specifications at this point is focused specifically on item type specification and how they align with the construct of academic literacy used for this test. The test developers turned to the specification of task types devised for TALL:

Table 1: Specifications and task types: TALL

| Specification (component of construct) | Task type(s) measuring / potentially measuring this component |
|--|--|
| Vocabulary comprehension | Vocabulary knowledge test, Longer reading passages, Text editing |
| Understanding metaphor & idiom | Longer reading passages |
| Textuality (cohesion and grammar) | Scrambled text, Text editing, (perhaps) Register and text type, Longer reading passages, Academic writing tasks |
| Understanding text type (genre) | Register and text type, Interpreting and understanding visual & graphic information, Scrambled text, Text editing, Longer reading passages, Academic writing tasks |
| Understanding visual & graphic information | Interpreting and understanding visual & graphic information, (potentially) Longer reading passages |
| Distinguishing essential/non-essential | Longer reading passages, Interpreting and understanding visual & graphic information, Academic writing tasks |

| Specification (component of construct) | Task type(s) measuring / potentially measuring this component |
|--|--|
| Numerical computation | Interpreting and understanding visual & graphic information, Longer reading passages |
| Extrapolation and application | Longer reading passages, Academic writing tasks, (potentially:) Interpreting and understanding visual & graphic information |
| Communicative function | Longer reading passages, (possibly also:) Text editing, Scrambled text |
| Making meaning beyond the sentence | Longer reading passages, Register and text type, Scrambled text, Interpreting and understanding visual & graphic information |

(Van Dyk & Weideman 2004:19)

With regards to TALPS, it was decided to include a section on argumentative writing. At postgraduate level it is essential that students follow specific academic writing conventions and it is important to test whether students were equipped with this knowledge. Butler (2009: 294) states: “In the development of TALPS we have also considered the importance of testing students’ productive writing ability specifically (in the production of an authentic academic text), as well as their editing ability”. In addition to the question on writing there is a question that tests students’ editing skills. Below is an outline by Butler (2009) of the eight sections that appear in TALPS. He has also given a brief explanation of what aspect of academic literacy each tests:

3.3 The eight subtests in TALPS

Section 1 of TALPS is a scrambled text in which sentences in a paragraph have been scrambled, and students have to rearrange the sentences so that the paragraph forms a cohesive whole. It therefore tests not only students’ ability in recognising text relations, drawing on their interpretative abilities regarding the context, but also their ability to recognise lexical clues contained in the sentences. Put differently: it assesses students’ command of various grammatical features of the text.

In Section 2, students’ knowledge of general academic vocabulary is assessed. The context created for this section is specifically that of the postgraduate academic environment, and the words tested are a selection of items from the different levels of the Coxhead academic word list (Coxhead, 2000).

Section 3 deals with visual and graphic literacy. Students are asked to interpret graphic information augmented by a short text discussion. This section mainly involves simple numerical computations and making inferences based on such calculations.

The **fourth section** emphasizes the importance of students being able to recognise different written text types. Students are requested to match two groups of sentences with regard to similarity in text type.

Section 5 includes a longer text that students have to read and subsequently answer comprehension type questions on the content of the text. Questions focus on students' abilities to classify and compare information, make inferences, recognise metaphorical language, recognise text relations and distinguish between essential and non-essential information.

Section 6 of the test assesses a number of academic literacy abilities. This question on grammar and text relations firstly provides students with a text they have to read where specific words have been omitted. Students then have to choose between 4 options regarding the place where these words have been left out in the sentences. The second part of the question requires that students, having been provided with the specific place where a word has been left out, choose between 4 options as to what is the correct word. The third part combines the formats of the first two in the sense that students are required to integrate the two tasks and do both simultaneously. They therefore have to find both the position where a word has been left out and the most suitable word that would fit that position. This section of the test assesses students' functional knowledge of sentence construction, word order, vocabulary, punctuation and at times communicative function (cf. Van Dyk & Weideman, 2004), with the main focus on the former, i.e. on grammatical or structural features of the language.

In **Section 7**, students' grammatical knowledge of English is assessed in the sense that they have to edit a short paragraph in which a number of typical language errors occur.

The last section of the test (**Section 8**) provides students with the opportunity to produce a written academic text. Similar to TALL, the reading texts selected for use in TALPS are topical in the sense that they all relate to the same topic. Students are then required to make use of any information in the test on the topic and write an argumentative text of approximately 300 words in which they present a structured argument. The argument is within the context of Africa. They also need to ensure that they give due recognition to the sources used in the test that they choose to include in their argument (they have to include a short list of at least 2 sources at the end of their texts). They further have to ensure that the text adheres to generally accepted academic writing conventions (such as formality of register, logical structure, acknowledging sources, etc.) (Butler 2009:294).

The next important step in the design process is to design appropriate task types in line with the blueprint. An important decision made by the test designers was to use a multiple choice format for the test, as had been done with TALL. The reasons for this, as outlined by Van Dyk and Weideman (2004), were the size of the population and the need to have the results ready urgently. The multiple choice format allowed for the test to be marked electronically rather than manually, thus ensuring that the results were ready on time. Using this format has, according to Van Dyk and Weideman, allowed them to become more “inventive and creative than we would otherwise have been, if we had simply succumbed to the prejudice that one cannot test (this or that aspect of) language in this way” (2004: 16). The following are examples of this inventiveness, based on a reading passage that was used to test the understanding of metaphor – a dimension of language use that conventionally might easily have been considered impossible to test in this format:

We should understand the phrase “milk in their blood” in the first sentence to mean that both men

- (a) have rare blood diseases inherited from their parents.
- (b) are soft-spoken, mild-mannered young farmers.
- (c) don't like to make profit at the expense of others.
- (d) are descended from a long line of dairy farmers.

Paragraph 2 speaks of ‘hatching a plan’. Normally, we would think of the thing that is hatched as

- (a) a door.
- (b) a loft.
- (c) a car.
- (d) an egg.

Or consider this one, which is designed to test the knowledge of the candidate regarding what counts as evidence:

In the second paragraph, we read that “milk farms have been the backbone of this country” for centuries. Which sentence in the fourth paragraph provides evidence of the claim that it has been so ‘for centuries’?

- (a) The first sentence
- (b) The second sentence

- (c) The third sentence
- (d) None of these

(Van Dyk & Weideman, 2004: 16).

Importantly, the authors of the article point out that the tasks the test takers are being asked to perform belong to a set of abilities or task types that are much broader in scope than that of a test that defines academic literacy in terms of skills, or reduces it to the mastery of sound, form and meaning (Van Dyk & Weideman, 2004: 16).

3.4 The process of development of TALPS

The draft process that the test developers followed in refining TALPS is easier understood in the form of a table:

Table 2: Table of subtests in drafts 1, 2 and final (TALPS)

| Task type | Marks (1st draft) | Marks (2nd draft) | Marks (3rd draft and first pilot) |
|-----------------------------|-------------------|-------------------|-----------------------------------|
| Scrambled text | 15 | 5 | 5 |
| Graphic and visual literacy | 16 | 16 | 10 |
| Dictionary definitions | 5 | 5 | |
| Academic vocabulary | 40 | 27 | 10 |
| Text type | 5 | 5 | 5 |
| Understanding texts | 60 | 60 | 25 |
| Grammar and text relations | 22 | 22 | 15 |
| Text edit | 10 | 10 | 10 |
| Total | 173 | 150 | 80 |

(Geldenhuys, 2007: 78)

The first draft of TALPS comprised of 173. The test was 150 minutes long and totalled 173 marks. At this stage the test did not include a question requiring students to write an argumentative text. The concern of the developers at this early stage was in writing the multiple-choice questions. These would then be analysed using TiaPlus Test and Item Analysis (Cito 2006) to determine which items did or did not test well. Items that did not test well were discarded. The second draft of TALPS totalled 150 marks and was 120 minutes long. The third draft version of the test became the first pilot for TALPS. Before the first pilot, by which time the test had been reduced to 100, items were evaluated by the designers to determine the appropriateness/strength of the item. Most changes were made in the Understanding text section. In the 100 item test this section had 45 items; in the 88 item test (which was the second pilot) it had 28 items. The final version of the test has 21 items in this section. Justification for this decision was drawn from the analyses done using the TiaPlus Test and Item Analysis Build 300 (Cito 2006). According to this, seven items had very high p-values, meaning that a high percentage of the test population got this answer correct.

4 Piloting the test

The test was piloted three times on different groups of students from different institutions. The test was first piloted on first year students who were registered for the then compulsory academic literacy course (EOT 110/120). Students were given one and a half hours to complete the 100 item test. It did not include the question requiring students to write an argumentative text.

The second pilot of TALPS was carried out on postgraduate students both at the University of Pretoria, and the University of Pretoria, in September 2007. This test comprised eighty-eight items and totalled 120 marks.

The third and final draft version of TALPS was made up of seventy-six items and eight sections, including the section requiring students to write a short argumentative text. This version of the test totalled 100 marks. The section on the Dictionary definitions was left out of this version of the test. The reason for this was that the descriptive statistics of the drafts of TALPS indicated that the Dictionary definitions question had a p-value of 84.2 for the 88 item pilot. Davies *et al.* (1999) explain that the higher the index, the easier the item. The closer the index is to 100% or 0%, the less differential information it can provide about candidates. They state that items that are excessively easy or very difficult are normally removed because they do not contribute to the test's discriminability (1999:95). The pilot for this version of the test was carried out in September 2007 on two groups of students: postgraduate students from Pretoria University and postgraduate students from the University of the Free State.

An analysis of the results of these pilots yields valuable information relating to the quality and efficiency of the test. Importantly, the primary purpose of such piloting is to construct an initial picture of test validity and reliability (Second Language Testing Inc. 2013). A study of the reliability and validity of TALPS is the focus of an earlier article

(See Rambiritch 2013). While the focus here is specifically to ensure transparency and accountability by telling the story of the design and development of TALPS, the story can never be complete without mention, again, of these issues. This will be done by looking specifically at the descriptive statistics of the TALPS third and final draft version.

Table 3: Descriptive statistics of the TALPS third and final draft version

| | TALPS COMBINED |
|--------------------------------------|----------------|
| N | 272 |
| Number of selected items | 76 |
| Minimum test score | 0 |
| Mean/average P- value | 51.88 (64.84) |
| Standard deviation | 13.32 |
| Cronbach's alpha | 0.92 |
| GLB | 0.99 |
| Standard error of measurement | 3.84 |
| Average <i>Rit</i> | 0.40 |

The instrument/package used (TiaPlus: cf. CITO, 2006) provides us with two measures: Cronbach's alpha and Greatest Lower Bound (GLB). All three pilots of the test have rendered very impressive reliability measures. The first pilot had a reliability of 0.85 (Cronbach's alpha) and 0.92 (GLB). The pre-final draft had measures of 0.93 (Cronbach's alpha) and 1.00 (GLB). The final version of the test had measures of 0.92 (Cronbach's alpha) and 0.99 (GLB), as indicated above. In terms of the standard error of measurement Kurpius and Stafford explain that a smaller standard error of measurement reflects a smaller error score and that the goal in reliability is to control error (2006:133). A higher reliability is therefore an indication of a small error of measurement. The 1st pilot of TALPS that has a reliability of 0.85, had a standard error of 4.30. When the reliability measures in subsequent pilots improved, the standard error of measurement dropped to 3.82 and 3.87. In the TALPS final version, the standard error of measurement for the combined groups of Pretoria University and the University of Pretoria, students is lower at 3.84.

One other statistical measure rendered by the package used is the average Rit-values or the discriminative ability of the test items. One of the main purposes of a test is to be able

to discriminate between the test-takers. According to Kurpius and Stafford (2006:115) a test cannot discriminate unless the items themselves discriminate between those who correctly answer the questions and those who do not. One of the main reasons to pilot a test is to determine which items discriminate well and which do not. The Rit-values for the 3rd pilot are relatively stable at 0.40, which is well above the 0.30 benchmark. In addition, the variance around the mean seems to be quite stable, suggesting a normal or even distribution of scores around the mean.

The brief discussion above on the reliability of TALPS is simply one indication of how test developers can make information available to the public. While the data, as it is presented above, may not mean anything to the layperson, proof that the test is reliable, that it is based on an appropriate construct, that each item was carefully designed or chosen, that the items were piloted and that weak items were discarded, all ensure that the test developers have considered and taken seriously the responsibility that comes with designing tests. Making this information available means that test takers are now equipped with information about the test and can now ask questions about the test. Test developers become real, not just experts 'hiding behind their designs'. Importantly, this kind of transparency and accountability ensures a channel of communication, not just between test developers and other experts in the field, but also between test developers and test takers.

5 Conclusion

The purpose of this article was specifically to tell the story of the design and development of the TALPS. The need to make this information available becomes relevant when one works within a framework that incorporates a concern for the empirical analyses of a test, as well as a concern for the social dimensions of language testing. As fair and responsible test developers it is our responsibility to ensure that all information about the test, its design and use, is freely available to those affected by or interested in its use. It should be the aim of test developers to design tests that are valid, reliable, accessible and transparent, by test developers who are willing to be accountable for their designs. Making information available is a first step in this direction.

References

- Bachman, L.F and Palmer, A.S. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Beu, D.S. and Buckley, M.R. 2004. Using accountability to create a more ethical climate. *Human resource management review* 14: 67-83.

- Bovens, M. 2005. Public accountability: A framework for the analysis and assessment of accountability arrangements in the public domain. In: Ferlie, E., Lynne, L. and Pollitt, C. (Eds.) 2005. *The Oxford handbook of public management*. Oxford: Oxford University Press. pp. 1-36.
- Butler, H.G. 2007. *A framework for course design in academic writing for tertiary education*. Unpublished doctoral thesis. Pretoria: University of Pretoria.
- Butler, H.G. 2009. The design of a postgraduate test of academic literacy: Accommodating student and supervisor perceptions. *Southern African linguistics and applied language studies* 27(3): 291-300.
- CITO. 2006. *TiaPlus, Classical test and item analysis* ©. Arnhem: Cito M. & R. Department.
- Council on Higher Education (CHE). 2004. *Higher education monitor: South African higher education: past, present and future*. Pretoria: Council on Higher Education.
- Coxhead, A. 2000. A new academic word list. *TESOL quarterly* 34(2): 213-238.
- Davidson, F. and Lynch, B.K. 2002. *Testcraft*. New Haven: Yale University Press.
- Davies, A., Brown, J.D., Elder, C., Hill, R.A., Lumley, T. and McNamara, T. (Eds.) 1999. *Studies in language testing: Dictionary of language testing*. Cambridge: Cambridge University Press.
- Fraser, N. 2008. Reframing justice in a globalising world. In: Olson, K. (Ed.) 2008. *Adding insult to injury: Nancy Fraser debates her critics*. London: Verso. pp. 273-291.
- Frink, D.D. and Klimoski, R.J. 2004. Advancing accountability theory and practice: Introduction to the human resource management review special edition. *Human resource management review* 14: 1-17.
- Fulcher, G. and Davidson, F. 2007. *Language testing and assessment: An advanced resource book*. New York: Routledge.
- Geldenhuys, J. 2007. Test efficiency and utility: Longer or shorter tests. *Ensovoort* 11 (2) : 71-82.
- Global Transparency Initiative. 2006. *Transparency charter for international financial institutions: Claiming our right to know*. [Online]. Available: www.ifitransparency.org. Date accessed: 2010/09/23.

- Kurpius, S.E.R. and Stafford, M.E. 2006. *Testing and measurement: A user-friendly guide*. California: Sage Publications.
- McNamara, T. and Roever, C. 2006. *Language testing: The social dimension*. USA: Blackwell Publishing.
- Mdepa, W. and Tshiwula, L. 2012. Student diversity in South African Higher Education. *Widening Participation and Lifelong Learning* 13: 19-33.
- Naurin, D. 2007. Transparency, publicity, accountability – the missing links. Unpublished paper delivered at the CONNEX-RG 2 workshop on 'Delegation and mechanisms of accountability in the EU', 8-9 March, Uppsala.
- Norton, B. 1997. Accountability in language assessment. In: Clapham, C. and Corson, D. (Eds.) 1997. *Language testing and assessment: Encyclopaedia of language and education* 7: 323-333. Dordrecht: Kluwer Academic.
- Rambiritch, A. 2013. Validating the Test of Academic Literacy for Postgraduate Students (TALPS). *Journal for Language Teaching* 47(1): 175-193.
- Schuurman, E. 2005. *The technological world picture and an ethics of responsibility: Struggles in the ethics of technology*. Sioux Center, Iowa: Dordt College Press.
- Second Language Testing Inc. 2013. *Pilot testing and field testing*. [Online]. Available: <http://2lti.com/test-development/pilot-testing-and-field-testing/>. Date accessed: 2013/12/10.
- Shohamy, E. 2001. *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
- Shohamy, E. 2008. Language policy and language assessment: The relationship. *Current issues in language planning* 9(3): 363-373.
- Sinclair, A. 1995. The chameleon of accountability: Forms and discourses. *Accounting, organisations and society* 20(2/3): 219-237.
- Transparency International. 2010. *What is transparency international?* [Online]. Available: <http://www.transparency.org/about us>. Date accessed: 2010/02/11.
- Van Dyk, T. and Weideman, A. 2004. Finding the right measure: From blueprint to specification to item type. *Journal for language teaching* 38(1): 15-24.
- Weideman, A. 2003. Assessing and developing academic literacy. *Per linguam* 19 (1 & 2): 55-65.

Weideman, A. 2006. Transparency and accountability in applied linguistics. *Southern African linguistics and applied language studies* 24(1): 71-86.

Weideman, A. 2009. Constitutive and regulative conditions for the assessment of academic literacy. *South African linguistics and applied language studies* 27(3): 235-251.

Weideman, A. and Van der Slik, F. 2008. The stability of test design: Measuring difference in performance across several administrations of a test of academic literacy. *Acta academica* 40 (1): 161-182.

ABOUT THE AUTHOR

Avasha Rambiritch

University of Pretoria, Hatfield Campus, Lynwood Road, Pretoria, 0002

Email: avasha.rambiritch@up.ac.za

The author is a lecturer at the University of Pretoria. Her teaching interests include academic writing at undergraduate and postgraduate level. She conducts research in the field of academic writing, tutor-assisted learning and language testing.