

2006. Appeared in *Language matters* 37 (1): 81-101.

ASSESSING ACADEMIC LITERACY IN A TASK-BASED APPROACH

Albert Weideman
Unit for Academic Literacy
University of Pretoria

Abstract

The *Test of academic literacy levels* (TALL) used by three South African universities (Pretoria, Stellenbosch and Northwest) provides a reliable and affordable alternative means of assessing the academic literacy of new entrants into the higher education sector. A close alignment is sought between the test, the task-based language instruction that follows its administration, and the learning and acquisition aimed for. The paper critically examines the construct of the test as well as its task types in light of various current discussions about authenticity. The paper concludes by suggesting a number of possible alternative task types that may achieve a closer alignment with the goals embodied in the construct. Various developmental, contextual, administrative and logistical constraints appear, however, to affect the level of resemblance to academic discourse of the test task types.

The assessment context

Every year the academic literacy of about 14000 new students enrolling at three South African universities, Northwest, Pretoria and Stellenbosch, is tested by means of a standardised assessment instrument, the *Test of academic literacy levels* (TALL). With some exceptions, students write the test in the language of their choice, i.e. either in Afrikaans (in which the test is called TAG, for *Toets vir akademiese geletterdheidsvlakke*), or in English.

The purpose of the test is to a large extent determined and required by the higher education context of South Africa, in which larger numbers of potentially underprepared students have found their way into tertiary studies. Increasingly, instead of blaming the schools for the degree of underpreparedness of students, universities and other institutions of higher education see this as a challenge, and to overcome the problem focus rather on intervention strategies, often in the form of academic literacy classes. While it is generally accepted that competence in the language or languages of instruction and learning is not the only factor that contributes to poor student performance, poor language ability is certainly viewed as a very important obstacle in this respect. Given this focus, it follows that tertiary institutions need a reliable measuring instrument to determine what level and nature of intervention is called for.

Of course, as national developments such as the benchmarking tests being prepared under the auspices of Higher Education South Africa (HESA) would

indicate, there is much more at stake than has been stated above. The National Benchmark Test project aims to give information on the ability of prospective students in language, mathematics and numeracy that will supplement the new Grade 12 exit exam results. Given the current pressures on universities, whose government funding is increasingly dependent on the successful completion of studies by students (“throughput”), there is a huge need to have a robust test of academic literacy that will also predict with some accuracy the potential of prospective students. These tests, however, are high stakes tests, that either exclude students from enrolling, or allow them into the academic fold.

Academic literacy tests may serve different purposes at institutions of higher education. The test currently under discussion, TALL, is not a high stakes test, since it is not an access test but a placement test, intended to channel students, if so required, into appropriate academic literacy support courses. The reason I mention the high stakes tests here in the discussion of what is a low to medium stakes test is that, since all of these tests are based on roughly the same construct or definition of academic literacy (see below), the current, more focussed discussion on one of that number may contribute also to their refinement and the critical reflection on them.

From both kinds of tests, access or high stakes tests, as well as placement or low to medium stakes tests, one would of course require that they do not have only reliability and validity, but that they meet a number of other criteria as well. It is the purpose of this article to articulate a selection of such criteria, in particular the notion of authenticity, and then discuss that in the context of its application to one test of academic literacy (TALL), and how that application then interacts with logistical, contextual and administrative constraints.

The merits of the TALL test construct and its specifications (Van Dyk & Weideman 2004a, 2004b), as well as its refinement (Van der Slik & Weideman 2005) and the strategies employed to ensure its transparency and accountability (Weideman 2003b, 2006) have been discussed in a number of other papers. The two outstanding features of the test are that it is efficient and reliable. The first, the efficiency of the test, which can be described here both as the economy of its administration or its logistical and administrative feasibility, is crucial to the current discussion: it takes only 60 minutes to complete. So large numbers of students can be tested at each institution on a single day, with the results becoming available within 18 hours. The second is that, for a low to medium stakes test, it has shown a remarkable degree of reliability. Table 1 below summarises its reliability indices across seven recent versions of the test (Cronbach’s α , as calculated by Itean; for other measures of reliability of the same test and their discussion, cf. Van der Slik & Weideman 2005: 27-30):

Date and version of the test	Alpha
2004 (University of Pretoria)	0.95
2005 (Northwest University)	0.94
2005 (University of Stellenbosch)	0.89
2005 (University of Pretoria)	0.93
2006 (University of Pretoria)	0.94
2006 (University of Stellenbosch)	0.91
2006 (Northwest University)	0.93
Average	0.93

Table 1: TALL: Reliability measures

The test construct

A test should not only be reliable, but also valid. In simple terms, this means that it should measure what it sets out to measure, and not something else. As Davies and Elder (2005: 797; cf. too McNamara 2005: 775) point out, the specificity of the context of tests, and with that the specificity of the discourse type assessed, has become increasingly important of late:

Much of the argument in language testing over the last period mirrors the argument in the wider social sciences and humanities area, that between the enlightenment (or universal) view that humanity (and experience) can be understood in similar ways and the relativist (or local) view that contexts are not just apparently but fundamentally different. This is the argument from postmodernism, which has insisted ... that it is unacceptable to assume that one size fits all.

It is not the purpose of this article to review the different kinds of validity distinguished in the literature. The Davies and Elder (2005) discussion referred to here provides ample information in this regard, and the application of such concepts to the test under discussion, TALL, is fully discussed in Weideman (2006). What concerns us, rather, is that the notion of validity today requires a contextual specificity. For testing of language ability within the context of an institution of higher education this means, in the first instance, that our starting point must be a credible definition of academic literacy. What we should be measuring should not be language ability in another context, say the ability to transact business by means of language, or the competence, for example, to use language for political, social or aesthetic purposes, but competence in handling academic language. Secondly, we must of course be able to operationalise our definition of this ability. If that is not feasible, no test will be possible. But that is a distinctly secondary consideration. The primary kind of validity that we seek in a

test of academic literacy levels is that it measures academic literacy according to a defensible definition.

The definition of academic literacy (Weideman 2003a: xi; cf. too Cliff, Yeld & Hanslo 2006, Cliff & Hanslo 2005) that lies at the basis of the test, and which the test designers claim derives from an open (interactive) rather than a restrictive (grammar-based) view of language, states that a student who is academically literate 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.

The test sets out to assess these ten components of academic literacy in a number of task types (cf. Van Dyk & Weideman 2004b) in seven different sections of the test. Though they may vary slightly from one version of the test to the next, these sections are:

- Section 1: *Scrambled text* (in which a scrambled paragraph is presented which students have to restore to its original order).
- Section 2: *Interpreting graphs and visual information* (which tests, among other things, the student's ability to interpret either a graph or a diagram, and to demonstrate a capacity for quantitative literacy [numeracy] related to academic tasks).
- Section 3: *Text type*. Here the students are presented with a number of sentences or phrases taken from a variety of text types or genres, which they have to match with a list of sentences or phrases from the same text types.
- Section 4: *Academic vocabulary*. Even though academic vocabulary is tested separately (and fairly traditionally) here, there are also vocabulary questions in some of the other sections.

- Section 5: *Understanding texts*. This section normally consists of one or more extended reading passage or passages, followed by questions focusing on critically important aspects of the construct, such as distinguishing between essential and non-essential information, or cause and effect, as well as inferencing, sequencing, defining, handling metaphor and idiom, and so forth.
- Section 6: *Text editing*. This part of the test, which relies on cloze procedure, normally has three sub-sections, though the text continues from the first to the last (for an example, see Unit for Academic Literacy 2006). In the first, a word is omitted, and students have to indicate the place where it is missing. In the second, the place where the missing word has been taken out is indicated, and students have to choose the appropriate word. In the third and final part, students have to indicate both place and missing word.
- Section 7: *Writing*. This section is used to test the ability of the student to make a short argument, which is normally connected to the theme of the text(s) that the student has read, as well as the topic of the scrambled paragraph and the text edit question. The test therefore contains (academic) information that is potentially useful in completing the writing section. It follows that this assignment has some relevance and authenticity, criteria that are important, as we shall note below, for task-based language assessment. For the purposes of this paper, however, the task in this section will not be considered, since this section is not always scored by the test administrators, who have a choice of marking it only in the case of, for example, borderline scores. Since such borderline cases are now being identified predominantly by empirical means (cf. Van der Slik & Weideman 2005), this choice is being exercised less and less by administrators.

The test is therefore fairly robust and stable not only in format, but also in that it is a reliable instrument, whose ongoing refinement is contributing to its validation and standardisation. As regards its validity, there has been no challenge to the empirical evidence regarding its discriminatory powers (cf. Van der Slik & Weideman 2005) and arguments thus far presented to academic audiences and in publications about the appropriateness of its construct (cf. Van Dyk & Weideman 2004a, Weideman 2006; also Cliff, Yeld & Hanslo 2006). It is, moreover, an affordable alternative, the costs of its development being contained by regular contributions in the form of expertise from each partnering institution, and agreements made annually in this regard. The major threat to its further development probably lies in an uncritical acceptance of its current form and content. The purpose of this paper is to see how, if we view the test critically from one vantage point, that of task-based language assessment, we can stimulate critical reflection on it.

The alignment of task-based instruction and assessment

For roughly 31% of the candidates who are annually assessed by TALL to be at risk as a result of too low a level of academic literacy, the University of Pretoria prescribes a compulsory intervention, in the form of an academic literacy course (Weideman 2003a). This intervention can best be described as a task-based course. It follows, in my opinion, that one should then enquire about the extent to which the initial assessment is itself task-based in nature. As one reviewer of an earlier version of this paper has pointed out, the idea of assessing of the efficacy of a task-based approach to teaching and learning *after* the intervention is relatively unproblematic. Why does the requirement of a task-based approach also affect the placement test that occurs before the intervention? The answer lies in the face validity of the test, in other words in the degree to which lay persons, such as the students who take the tests, and perhaps their parents, perceive the test to be a credible and relevant instrument with which to assess academic language ability. We take great care to ensure that the intervention that follows is perceived to be appropriate and relevant. We should do the same, I would argue, for the initial channelling mechanism, the test. So the question is: how relevant to a task-based approach to language learning and teaching are the various task types that form part of the test described above? Are they, too, embedded within a broadly defined task-based approach?

If, as Brindley (2002: 465) implies, we are entering a phase in language testing where we seek “a closer alignment between assessment and learning”, this would also apply to the alignment of both of these with instructional approach. Such potentially positive “washback” (Brindley 2002: 467) among assessment, instruction and learning is the focus of this paper.

While there may still be some debate about our ability to make language learning claims for a task-based approach to academic literacy, a lack of alignment between, for example, testing and teaching would certainly be open to valid criticism. If the degree of alignment between assessment and teaching is not high, it would justifiably give rise to concern.

From a test designer’s point of view, the question therefore would be: what would a task-based approach to language assessment, instruction and learning have to contribute (additionally) to the current format and task types within TALL? What further requirements would such an approach generate for an already stable, valid and reliable language assessment instrument?

Characterisations of a task-based approach

To see what such additional requirements might be, one may look at a number of general characterisations of a task-based approach. So, for example, one may agree

with Littlewood's (2004: 319) recent assessment that task-based language teaching "has achieved something of the status of a new orthodoxy." Similarly, there is a close relationship between task-based teaching and the other contender for this status, communicative language teaching (CLT); I fully agree with Wesche & Skehan's (2002: 228) recent conclusion that task-based language teaching "could be considered simply a more thoroughgoing version of CLT."

CLT in one of its earliest forms was of course concerned, like task-based language teaching, with authenticity, "exposing students to the language of the real world" (Guariento & Morley 2001: 347). The difference, as these authors also point out, is that those early concerns with authenticity were focused more on texts than on tasks. Nonetheless, in broad terms it has been a shared concern.

There is little doubt that a task-based approach to language teaching is conceptually closely linked to notions of authenticity. This is not to say, of course, that either the link or the notion itself is free from controversy. There is a persistent hedging in the literature whenever the idea of authenticity is discussed, with a retreat into a relatively safe and sometimes necessarily bland definition, as when defining tasks used in language teaching and learning as relating to things people do in "the real world", i.e. using language such as that we use in doing everyday things, which is an approximation of, or bears at least a resemblance to the kind of language expected there (Skehan 2003: 3). But there is also another interpretation of the notion of authenticity, and both this and the previous interpretation are relevant to the design of task-based language assessment. This interpretation is in evidence, for example, in Guariento and Morley's (2001: 349) observation that authenticity "lies not only in the 'genuineness' of text, but has much to do with the notion of task" itself. That is, the subjective engagement by learners with the task (Guariento & Morley 2001: 350) authenticates it when it awakens their interest and an understanding of its relevance. We return below to a discussion of the various forms of authenticity that are relevant to the assessment of language ability in a task-based approach.

Finally, most characterisations of task-based teaching rely on what is now an almost standard definition, summarised in Wesche & Skehan's (2002: 217) recent survey. In this definition, a task is an activity promoting language learning

- in which meaning is primary, and
- through which communication must take place (as in regular CLT definitions);
- that has a link with real-world activities;
- and its completion has priority, *i.a.* because
- the success of its completion is assessed by outcome(s).

These characteristics of a task, especially the third, its link with real-world activities, are equally relevant to a task-based approach to language assessment or testing. In conceptions of task-based language assessment we indeed find the same

conceptual link to the notion of authenticity. This is particularly pertinent in the case of testing for academic literacy, which entails the assessment of language used for a specific purpose. In fact, as Lewkowicz (2000: 44) remarks: “Authenticity is ... one of two features which distinguishes such tests from more general purpose tests of language” ability. Therefore, as in the case of language teaching, task-based language testing may require test tasks to simulate or mirror real-life tasks (Lewkowicz 2000: 47). Similarly, as in the case of discussions focussed more on task-based language learning and teaching, we find an interpretation of authenticity in the discussions on task-based language assessment that highlights the subjective engagement of the test taker with the task as that which validates and therefore authenticates it. In this view authenticity is “a quality arising from the test takers’ involvement in test tasks” (Lewkowicz 2000: 48).

Given the prominence of the notions of resemblance to real language use and user engagement in the characterisations we have considered so far, it may be worthwhile to focus our attention for a moment on these in the next section, before proceeding to ask how such notions may be useful to those designing language teaching tests or assessment instruments.

Four (three) kinds of authenticity

To the language test designer, Guariento and Morley’s (2001) discussion of different kinds of authenticity is probably the most immediately useful. They distinguish between four different kinds of authenticity (which may actually be only three, since one appears to be a special case of another). The summary below is my interpretation of their distinctions (2001: 349-351):

- The first kind of authenticity is achieved through genuinely purposeful communication, i.e. when the communication really involves a sharing and transfer of information, and/or negotiation of meaning, and takes place in the context of real-time interaction to achieve its communicative goal.
- The second kind is accomplished when the task relates to “real-world” needs.
- The third type of authenticity occurs in the language learning situation, provided by the context of the classroom.
- The fourth kind of authenticity is realised in the subjective engagement of the learner with the task, involving them to such a degree that they may even be asked to select the tasks themselves.

If the argument is acceptable that the third (authenticity achieved through genuine communication within the classroom) is a special case of the first kind, one may observe further that both one and three in any event echo exactly the requirements

of any communicative task, in any interpretation of CLT (i.e. the ‘authentic text’ direction of first generation CLT [cf. Lewkowicz 2000: 43], or mainstream, functional interpretations, or even later interpretations of a ‘humanistic’ or ‘natural’ nature). What all of these different interpretations of CLT have in common is that in them a language learning task always and of necessity has to involve communication across an information gap, i.e. be real and purposeful communication (for discussion, cf. Roberts 1982, Weideman 2002: 31-33). In this sense both the third and first kinds of authenticity are no different in task-based teaching than they would have been in any teaching and learning organised around communicative principles, and in this respect therefore have nothing more to add to a test that already takes an interactive, open or communicative view of language.

Within these four parameters, then, the two that could potentially add further value to a test of language ability are those that relate to the real-world needs or tasks (the second kind) and to the level of engagement of the test taker with the task (type four). I would agree that both of the two remaining factors (the relation of a task to real-world needs, and the engagement of the test taker with the task) are perhaps still not fully articulated in the definitions above, and retain a certain conceptual vagueness. The best interpretations seem to be that they refer to the subjective judgements that one might make regarding, respectively, the resemblance of a task to one that occurs or may occur in the context of academic work, and the way that the task requires that the test taker focus what are called “attentional resources” (Elder & Iwashita 2005; also Tavakoli & Skehan 2005) on it. Thus a task may be more complex in that it involves more cognitive processes than another. As we will see below, a sub-test of academic vocabulary that requires the test taker to place a single word (out of a choice of four) into a gap left in a sentence that resembles academic discourse, necessarily requires less engagement from the test taker than another, such as the text editing question (cf. Van Dyk & Weideman 2004b, and the sample test under Unit for Academic Literacy 2006) that requires test takers to consider two sets of choices, and two sets of gaps.

Assessing authenticity

Having now isolated two aspects of the idea of authenticity as potentially useful, it remains for us to indicate how the currently available task types in the test under discussion fare when measured against these. The analysis can best be summarised in the form of a matrix that measures both aspects, the level of engagement of the test taker with the task, as defined above, and the relation to the “real world”, or the resemblance that the task has to actually or potentially occurring academic language:

engagement	strong	Not ideal, but passable	Best option
	weak	Not aligned with goal	Not ideal, but passable
		weak	strong

relation to “real world”

Figure 1: Matrix for assessing authenticity

Eliminating the last section of the test (section 7) for the reasons discussed above, and taking into consideration only the first six, i.e. those that are scored objectively and are presented to the test takers in multiple choice format, we may try to gauge the extent to which they fulfil the criteria for authenticity (weak or strong engagement, or weak or strong relation to real world tasks) as in the matrix above.

The outcome of such an analysis as presented below is perhaps not entirely uncontroversial, but is based on a good deal of involvement with the design of the test. If anything, it may be on the conservative side, i.e. many of those who have been involved in developing the test may tend to judge the different task types more leniently. The reason for adopting a more severe assessment is to get as clear an indication as possible, of course, of where the test may be improved. This analysis is presented below in Figure 2:

engagement	strong	Scrambled text Text type Text editing	Interpreting graphs and visual information Understanding texts
	weak	Academic vocabulary	
		weak	strong

relation to “real world”

Figure 2: Matrix for assessing authenticity

Many would argue that in this analysis, the testing of vocabulary (in the conventional form of a sentence, but employing Coxhead’s [2000] Academic word list) may be too harshly adjudged to be weak, especially in terms of test taker engagement. The desirable task types, however, are few: only two, dealing with the

interpretation of graphs and visual information, and understanding texts, are in the ideal quadrant, since they both bear a strong relation to real academic tasks, and at the same time strongly engage test takers. While not entirely undesirable, the three task types in the moderately authentic category (since they at least strongly engage test takers) may, like the vocabulary testing, either be improved or replaced by other task types, which would fare better in this kind of analysis.

In the next section, we present and discuss the merits of several examples of task types that may either replace, or perhaps supplement, the existing ones.

Some alternative task types

This section contains examples of three alternative task types. They involve first a writing task (the questions on “References”), then, second, a combination of a listening comprehension and writing task that bears some resemblance to note taking activity in an academic classroom (“Listening and taking notes”), and third a real-time, real-life survey of the way in which fellow students waste time (“Doing a survey”).

References

Imagine that you have gone to the library to search for information in the form of books, articles and other material, on the topic of ‘Making effective presentations’. You have found a number of possible sources, and have made notes from all of them for use in your assignment on this topic, but have not had the time to arrange them in proper alphabetical and chronological sequence.

Look at your notes below, then place the entry for each source in the correct order, as for a bibliography, in the space provided below:

- (a) Jay, R. 2000. *How to write proposals and reports that get results*. London: Pitman.
- (b) Dickinson, S. *Effective presentation*. 1998. London: Orion Business.
- (c) Hager, P.J., H.J. Scheiber & N.C. Corbin. 1997. *Designing and delivering scientific, technical, and managerial presentations*. New York: Wiley-Interscience.
- (d) Chemical and Process Engineering, University of Newcastle-upon-Tyne. 2001. Presentation skills. Available <http://lorien.ncl.ac.uk/ming/Dept/Tips/present/present.htm>.
- (e) Jay, R. & A. Jay. 1994. *Effective presentation: powerful ways to make your presentations more effective*. Prentice-Hall: London.

- The entry I placed **first** is (a) (b) (c) (d) (e)
 The entry I placed **second** is (a) (b) (c) (d) (e)
 The entry I placed **third** is (a) (b) (c) (d) (e)
 The entry I placed **fourth** is (a) (b) (c) (d) (e)
 The entry I placed **fifth** is (a) (b) (c) (d) (e)

The entry with the **date** of publication in the wrong place is (b) (c) (d) (e)

The entry that has the **place** of publication in the wrong place is

- (a) Jay (2000)
- (b) Hager *et al.* (1997)
- (c) Chemical and Process ... (2001)
- (d) Jay & Jay (1994)

The relationship of this kind of task with real academic tasks is probably beyond doubt, and the task would probably score high on engagement as well. However, while knowledge of how to alphabetise and order a set of references is part of academic life, is it fair to expect academic beginners, the population that the test is aimed at, to know this even before they commence their studies? The task type scores high, but other considerations may therefore militate against using it.

The text used in the second example below, on note taking (listening and writing), is well within the standards set for reading ease and grade level that the texts used in TALL are required to conform to. Its grade level (Table 2, below), for example, is well below grade 10, where texts intended for reading in TALL normally are selected from examples that lie between 10.5 and 11.5. The readability statistics yielded by Microsoft's Word programme are used here, simply because they provide all test developers, sometimes working in diverse locations, with a single, uniform measure of reading ease (Flesch) and grade level (Flesch-Kincaid). Furthermore, the Flesch Reading Ease level of this particular test (at 58.1) also comes out above the 45 to 55 level usually required of reading texts in TALL. The text is therefore easier in a number of respects than those used for reading, and this makes sense: it is intended to be used as a listening test.

Readability Statistics	
Counts	
Words	238
Characters	1096
Paragraphs	4
Sentences	13
Averages	
Sentences per Paragraph	4.3
Words per Sentence	18.0
Characters per Word	4.4
Readability	
Passive Sentences	23%
Flesch Reading Ease	58.1
Flesch-Kincaid Grade Level	9.6

Table 2: Readability statistics: listening and note taking text

Here is the second example, adapted from Weideman (2003a: 16-19):

Listening and taking notes

Read the following to students, who have in front of them the outline of the mind map that they have to complete:

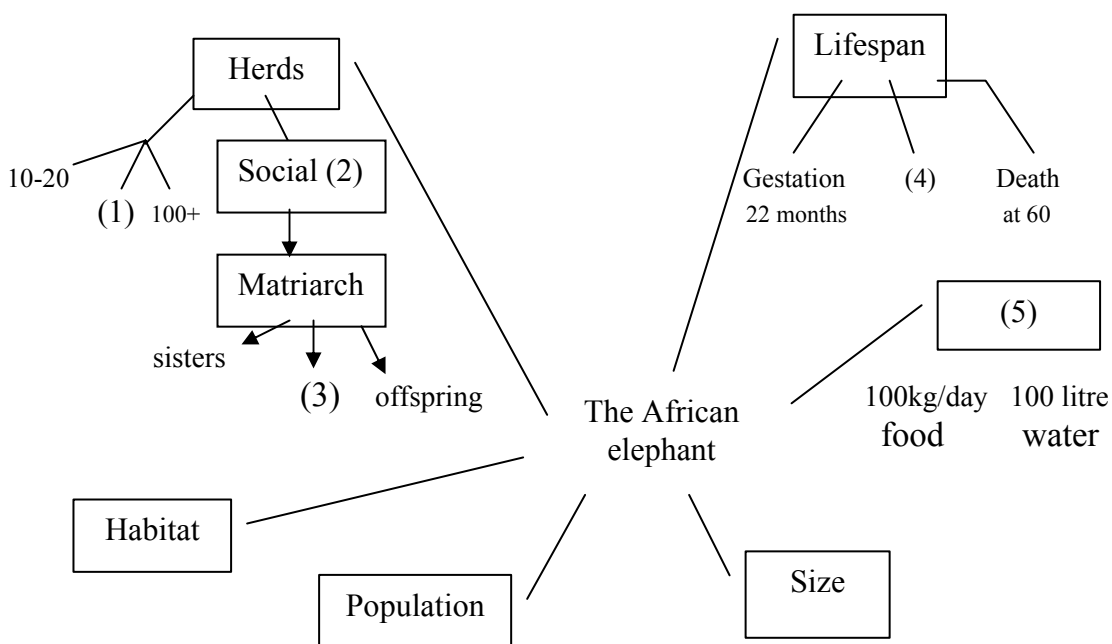
The African elephant

Elephants essentially live in herds and may be found in groups of anything between 10 and 20 or up to 50 and more, and, in rare cases, in excess of 100. Their highly developed social structure, however, remains consistent throughout. Family units are led by a cow elephant, or matriarch, and a typical family herd consists of cow elephants of various ages: the leader, and her sisters, their daughters, and their offspring.

The lifespan of an elephant is long and often eventful. For one thing, elephants have a long gestation period: 22 months. Newborn calves weigh some 120 kilograms, yet, in spite of their size, fit easily under their mother's belly, where they suckle. When a bull reaches puberty at between 12 and 14 years of age, his time within the herd comes to an end, although he may return or join other herds for short periods. As an adult, right up to his death at any time up to age 60, he will not remain with any family herd permanently.

The elephant's diet is equally remarkable. The average elephant probably consumes well over 100 kilograms of food per day. They forage with their trunks, through which they also draw up water. As much as 100 litres may be consumed at a time. Elephants spend between 16 and 18 hours a day eating both grass – which makes up some 80% of their diet – and browse.

Notes taken in a linear way (writing from left to right, top to bottom) are sometimes not as effective as those done in the form of a mind map. Your task is to complete the mind map below with information that your lecturer will read to you. Select the elements for completing the mind map from the choices given below:



- (1) The most appropriate choice here is
- (a) between 30 and 40
 - (b) more than 40
 - (c) more than 50
 - (d) about 70 or 70

- (2) The term that is used here is
- (a) development
 - (b) structure
 - (c) herd
 - (d) family

- (3) The word that fits here is
- (a) elephants
 - (b) age groups
 - (c) children
 - (d) daughters

- (4) The phrase that most appropriately completes this is
- (a) phases of life of the African elephant
 - (b) bulls leave herd at puberty (12-14 years)
 - (c) cows live in herds till their death
 - (d) the lifespan is long and eventful
- (5) The word that fits best here is
- (a) Habitat
 - (b) Numbers
 - (c) Size
 - (d) Diet

Again, this task will score high both on test taker engagement and on the way that it mirrors real academic tasks. It would, however, require (a) fairly sophisticated sound equipment, especially when administered (as TALL is) to large groups of up to 600 test takers in a single venue at a time; (b) a duplication of this equipment in every venue where the test is being administered simultaneously (i.e., between four and nine sets of such apparatus on each of seven campuses); and (c) a thorough process of standardisation, e.g. in using a specific (South African, Black South African or RP) accent to read the text. All of these points are exactly where the commercially available test that preceded TALL failed: since it required such sound equipment, the logistical and administrative trouble to get that right simply was not worth the effort.

This indicates that, though the task type may be highly relevant and engaging, its use may have to be restricted to, say, second chance testing or smaller groups. Indeed, these kinds of smaller groups of test takers exist, in the administration of a test for latecomers, and in the retesting (second chance testing) of borderline cases. It may therefore have limited use, and can easily be adapted for use in a computer laboratory, thus accommodating a few hundred test takers simultaneously.

The third example is probably the most authentic of them all:

Doing a survey

Most students complain that they procrastinate, and therefore have difficulties in disciplining themselves to study. They waste time on activities that could have been spent studying, and regret that afterwards. You may have had the same experience when you were preparing for an important school exam.

In this task, you have to do a survey among your fellow students about the time they waste. First complete the questionnaire below for yourself. Calculate the time you waste according to the percentage of the full normal working day the wasteful activity takes up. Here is a table that converts time to approximate percentages:

Time	% of day
½ hour	2%
1 hour	4%
2 hours	8%
3 hours	13%
4 hours	17%
5 hours	21%

You have 5 minutes to complete this task.

Activities on which I waste time (as percentage of a normal working day)

	% of day
Playing pool	
Playing TV games	
Watching late afternoon / early evening TV shows (17:00 - 19:00)	
Not doing things right the first time, thus having to repeat them	
Unrestricted or unplanned socialising	

After you have completed the questionnaire for yourself, you will be given the opportunity to find out from five other students what their answers look like. Collect and then summarise the information in the table below, and calculate the average for every category:

	Student 1	Student 2	Student 3	Student 4	Student 5	Averages
Playing pool						
TV games						
Watching soaps						
Repeating things						
Unplanned socialising						

A student who has done a similar survey to the one you have done has come up with the following data:

Time wasted by a sample group of undergraduate students (expressed as % of day)

	Student 1	Student 2	Student 3	Student 4	Student 5	Averages
Playing pool	0%	2%	0%	4%	0%	(1)
TV games	2%	8%	13%	0%	17%	(2)
Watching soaps	0%	2%	2%	2%	(3)	2%
Repeating things	4%	0%	8%	4%	4%	4%
Unplanned socialising	17%	8%	10%	13%	2%	(4)
<i>Totals</i>	23%	20%	(5)	(6)	27%	

Answer the following by referring to the data in the table above:

- (1) The average time wasted on this kind of activity is
 - (a) 1.0%
 - (b) 1.2%
 - (c) 1.3%
 - (d) 1.5%

- (2) The average time wasted in this group on playing TV games is
 - (a) 5%
 - (b) 6%
 - (c) 7%
 - (d) 8%

- (3) The percentage of the day wasted by student 5 in the survey watching TV soaps is
 - (a) 4%
 - (b) 6%
 - (c) 8%
 - (d) 10%

- (4) The average percentage of time wasted on unplanned socialising is
 - (a) 8%
 - (b) 9%
 - (c) 10%
 - (d) 11%
 - (e) 12%

- (5) The total time wasted by student 3 adds up to
 - (a) 30%
 - (b) 31%
 - (c) 32%
 - (d) 33%

- (6) Student 4 wastes an average equal to that wasted by student
- (a) 1
 - (b) 2
 - (c) 3
 - (d) 5
- (7) The student with the worst time wasting problem is student number
- (a) 1
 - (b) 3
 - (c) 4
 - (d) 5
- (8) The evidence for my answer to the previous question can be found in the space in the table above marked
- (a) (6)
 - (b) (5)
 - (c) (4)
 - (d) (2)
- (9) The student with the least of a time-wasting problem is student number
- (a) 1
 - (b) 2
 - (c) 3
 - (d) 4
- (10) The category that needs most attention if students want to stop wasting time is
- (a) Unplanned socialising
 - (b) Having to do things over
 - (c) Playing TV games
 - (d) Playing other games

Now compare your findings, or your individual information, with the data of the other survey, above. Write at least five sentences, which (a) should be free from spelling and punctuation errors, (b) contain no grammatical mistakes, and (c) clearly state what the comparison is.

Surveying, summarising, comparing: all of these are highly relevant academic tasks, and especially so in terms of the construct of the test under discussion. Again, this kind of task should score high both in respect of test taker engagement and its resemblance to real life academic tasks. But once again, with up to 8000 students writing the test in a single day, and the availability of venues appropriate to the kind of activity implied here, its practical administration looks like a virtual impossibility. It looks more like a typical classroom activity, and would certainly be more appropriate, and less of a logistical and administrative nightmare, in such a context.

Conclusion

Perhaps, in the three examples noted above, however laudable they would be in respect of the two dimensions of authenticity under discussion here, the initial lesson seems to be that their adoption will not be unproblematic. Thus, however useful and desirable they may be as authentic task types that strongly engage test takers and have a high degree of actuality and relevance, there are logistical, contextual and administrative constraints that either prevent them from being used, or would limit their use to smaller groups than those currently being targeted, at least in the first administration of the test. They will also not be as economical with time as the current test, which will further undermine their usefulness.

Of course, one may argue that three examples are hardly enough evidence for abandoning the project of appropriately aligning testing with teaching and learning. One should indeed keep on looking for alternatives, and attempt to deal in the test design with the logistical and administrative constraints that more desirable task types throw up. But it is sobering to note, with Lewkowicz (2000: 50), that

... test development is an evolutionary process during which changes and modifications are likely to be continually introduced. Such changes may, ultimately ... affect the degree of correspondence between the test tasks and TLU (target language use) tasks.

If time and resources were no consideration, a whole new set of more desirable alternative task types might well have been adopted for TALL. For the moment, however, it would appear that we have to tolerate less than perfect correspondence with real academic tasks in this particular test of academic literacy, and that we must seek authenticity more in that dimension of it that emphasises the test takers' engagement with the tasks that it contains.

The discussion here does not yet provide a larger framework in which to view task-based test types and items, especially as these function in the placement and access tests mentioned at the beginning. What we do know, however, is that some of the existing task types that can be identified by applying the criteria from Figure 1 as "Best options" (cf. Figure 2), viz. *Interpreting graphs and visual information* and *Understanding texts*, are already not only valid, but also highly reliable item types (cf., e.g., the findings in Van Dyk & Weideman 2004b, and Van der Slik & Weideman 2005). The challenge will be to design further task-based test items that score as high on the criteria for authenticity (level of engagement and resemblance to real discourse) used here, and that are also feasible, with strong discriminatory and predictive powers. This kind of initiative is most likely to yield results if it is tackled as a larger scale research project.

References

- Brindley, G. 2002. Issues in language assessment. In R.B. Kaplan (ed.) 2002: 459-470.
- Cliff, A.F. & M. Hanslo. 2005. The use of 'alternate' assessments as contributors to processes for selecting applicants to health sciences' faculties. Paper read at the Europe Conference for Medical Education, Amsterdam.
- Cliff, A.F., N. Yeld & M. Hanslo. 2006. Assessing the academic literacy skills of entry-level students, using the Placement Test in English for Educational Purposes (PTEEP). Forthcoming in *Assessment in education*.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34 (2): 213-238.
- Davies, A. & C. Elder. 2005. Validity and validation in language testing. In E. Hinkel (ed.) 2005: 795-813.
- Elder, C. & N. Iwashita. 2005. Planning for test performance. Does it make a difference? In R. Ellis (ed.) 2005: 219-238.
- Ellis, R. 2005. *Planning and task performance in a second language*. Amsterdam: John Benjamins.
- Guariento, W. & J. Morley. 2001. Text and task authenticity in the EFL classroom. *ELT Journal* 55 (4): 347-353.
- Hinkel, E. (ed.). 2005. *Handbook of research in second language teaching and learning*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kaplan, R.B. (ed.). 2002. *The Oxford handbook of applied linguistics*. Oxford: Oxford University Press.
- Lewkowicz, J.A. 2000. Authenticity in language testing: some outstanding questions. *Language testing* 17 (1): 43-64.
- Littlewood, W. 2004. The task-based approach: some questions and suggestions. *ELT Journal* 58 (4): 319-326.
- McNamara, T. 2005. Introduction to Section VI: Second language testing and assessment. In E. Hinkel (ed.) 2005: 775-778.
- Roberts, J.T. 1982. Recent developments in ELT, Parts I & II. *Language teaching* 15 (2 & 3): 94-110; 174-194.
- Skehan, P. 2003. Task-based instruction. *Language teaching* 36: 1-14.
- Tavakoli, P. & P. Skehan. 2005. Strategic planning, task structure and performance testing. In R. Ellis (ed.) 2005: 239-273.
- Unit for Academic Literacy. 2006. Academic literacy and sample test. [Online]. Available <http://www.up.ac.za/academic/humanities/eng/eng/unitlangskills/eng/fac.htm>. Accessed 14 April 2005.
- Van der Slik, F. & A. Weideman. 2005. The refinement of a test of academic literacy. *Per linguam* 21 (1): 23-35.

- Van Dyk, T. & A. Weideman. 2004a. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *SAALT Journal for language teaching* 38 (1): 1-13.
- Van Dyk, T. & A. Weideman. 2004b. Finding the right measure: from blueprint to specification to item type. *SAALT Journal for language teaching*. 38 (1): 15-24.
- Weideman, A.J. 2002. *Designing language teaching: on becoming a reflective professional*. Pretoria: BE at UP.
- Weideman, A.J. 2003a. *Academic literacy: prepare to learn*. Pretoria: Van Schaik.
- Weideman, A.J. 2003b. Assessing and developing academic literacy. *Per linguam* 19 (1 & 2): 55-65.
- Weideman, A.J. 2006. Transparency and accountability in applied linguistics. Forthcoming in *Southern African linguistics and applied language studies* 24 (1).
- Wesche, M.B. & P. Skehan. 2002. Communicative, task-based, and content-based language instruction. In R.B. Kaplan (ed.). 2002: 207-228.