The Multi-Divisible or Unitary Nature of Reading: The language tester between Scylla and Charybdis

Cyril J. Weir and Don Porter
CALS, University of Reading

This paper discusses the relevance for the valid testing of reading of the difference between a 'unitary skill' approach to reading, i.e. seeing reading as a single, undifferentiated skills activity, and a multiskills approach, seeing the reading activity as composed of a number of distinguishable skills. It reviews some of the evidence, both quantitative and qualitative, in favour of each contrasting view, and comments on the potential dangers for valid testing of the adoption of one or other approach. It produces experimental evidence from different areas, and using different test methods, which suggests that a distinction may be drawn between language-based skills and 'global' reading skills, and comments on the implications of this for EAP testing of reading.

INTRODUCTION

Employers, university admissions officers, teachers and other users of test data often need specific and reliable information about a candidate's reading ability. In the attempt to satisfy that need, the language tester has to try to reduce the possibility of muddled measurement by eliminating the influence of irrelevant factors associated with the test method, and by focusing on a maximally clear characterization of the construct of reading per se. The tester might therefore avoid tasks such as selective summary based on prior reading of texts where the extended writing involved in task completion might interfere with any inferences we might wish to make concerning candidates' reading abilities alone. But first and foremost, the tester is obliged to be as explicit as possible concerning the nature of the ability about which the test is designed to provide information.

Williams and Moran (1989) give a representative account of the current consensus among writers of teaching materials on the nature of reading comprehension, stating that:

"While materials writers may disagree on the emphasis to be devoted to any particular skill, there seems to be substantial agreement on the importance of such skills as guessing the meaning of unknown words, identifying anaphoric reference, identifying the main idea, and inference." (p.224).

Cyril WEIR is Associate Director of the Testing and Evaluation Unit, Centre for Applied Language Studies, University of Reading. He is the author of Communicative Language Testing Understanding and Developing Language Tests, and a co-author of Evaluation in ELT.

Don PORTER is Director of the Testing and Evaluation Unit. He has published widely on language testing issues, and has been involved in the development of many language tests in Britain and overseas.

The authors may be contacted via Cyril Weir, CALS, University of Reading, P.O. Box 16 Whiteknights, Reading RG6 2AA.
In a less pedagogically oriented discussion, Grabe (1991:379-382) may also be seen as representative of current views. He comments on the importance of automaticity in reading, particularly in word identification, and also identifies as components of skilled reading: syntactic knowledge; knowledge of formal discourse structure (formal schemata); content and background knowledge (content schemata); and metacognitive knowledge and skill monitoring, e.g. recognising the more important information in a text, skimming, and searching for specific information. Grabe (1991:382) concludes:

"A 'reading components' perspective is an appropriate research direction to the extent that such an approach leads to important insights into the reading process. In this respect, it... is indeed a useful approach".

Such a focus would then seem to accept that reading can be broken down into "underlying skills components" for the purposes of teaching and testing — a view shared by many language teachers and testers.

### Table 1: Summary checklist of operations in reading (Weir, 1993:73)

<table>
<thead>
<tr>
<th>Reading a text quickly:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying: which part, or whether part or whole, is relevant to an established need.</td>
</tr>
<tr>
<td>Surveying for gist.</td>
</tr>
<tr>
<td>Scanning to locate specific information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reading carefully to understand main ideas and important detail:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading for main ideas and important detail (might include tracing the development of an argument, reducing what is read to an outline of the main points and important details).</td>
</tr>
<tr>
<td>Distinguishing fact from opinion, recognising attitude to reader and topic (e.g. persuasion or explanation).</td>
</tr>
<tr>
<td>Understanding inferred meaning.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At some level all of the above might include a knowledge of the following more specifically linguistic contributory skills:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding concepts (grammatical notions) such as: cause, result, purpose, comparison.</td>
</tr>
<tr>
<td>Understanding syntactic structure of sentence and clause.</td>
</tr>
<tr>
<td>Understanding discourse markers.</td>
</tr>
<tr>
<td>Understanding lexical and/or grammatical cohesion.</td>
</tr>
<tr>
<td>Understanding lexis; including deducing meaning of lexical items from morphology and context.</td>
</tr>
</tbody>
</table>

A recent three-level version of such a breakdown is provided in Table 1 above, taken from Weir, 1993. The reader will note that Williams and Moran (1989) and Grabe (1991) both identify reading skill components from all three of the levels A, B and C in Table 1. Their inclusion of word-level components relating to more specifically linguistic comprehension is indicative that this is seen by many people as an important part of reading, not as something separate. The contribution of the latter to tests of reading ability is the major issue which this paper will address.

## MULTI-DIVISIBILITY IN CURRENT PRACTICE: ON A WING AND A PRAYER?

Just as a current consensus holds that there exist systematic differences between the skills of listening, reading, writing and speaking, so too it is held by many teachers, textbook writers, and the constructors of language tests that in reading itself there are different skill components, and that students may exhibit differences in level of proficiency across these. It is moreover often claimed by practitioners that sets of reading skill components provide useful frameworks on which to base course design, teaching, and test and materials development (see Lumley, 1993).

It is widespread practice among test-developers to focus on these reading skill components, either singly or in combination, when constructing test items relating to a text, even though the sum of these parts — the answers to these test items — might not necessarily equate fully with what the reader would normally take away from the text. Indeed, whatever theoretical position the test developer takes, the need to construct individual test items will exert strong pressure to attempt to measure individual reading skill components and strategies, or combinations of them. If specific components or strategies could be clearly identified as making an important contribution to the reading process, then it would of course be at least possible, if not necessary, to test these and to use the composite results for reporting on the reading proficiency revealed. It would moreover be necessary, once hypothesized skill components and strategies were substantiated, to determine the nature and strength of interactions between them. It is however as yet too early to explore such interactions; thus far we can have but little idea whether any skill components may be superordinate to others, or the extent to which "higher order" operations within reading depend on "lower order" operations.

Despite the wide-spread influence of a multi-divisible view of reading on current practice, caution needs to be exercised. Acceptance of such a view by practitioners tends to be anecdotal, and based on pedagogical "experience". What does the research show?
CASTING DOUBT ON MULTI-DIVISIBILITY

In opposition to a multi-divisible view of reading, a substantial number of studies have found that it is not possible to differentiate between reading skill components, either through empirical demonstration of the separate functioning of such components when these are operationalised in language test items, or through the judgement of experts on what the focus of such test items actually is (see e.g. Alderson 1990a, Alderson and Lukmani 1989, Carver 1992, Rosenshine 1980, Rost 1993).

Quantitative research

A number of empirical studies, typically using factor analysis, have cast doubt on the multi-divisible nature of reading e.g. Lunzer et al. 1979, Rosenshine 1980 and Rost 1993). Factor analysis is a statistical procedure for extracting the extent to which putatively different variables – in our case the so-called “subskills” in reading – in fact function in a similar manner. If a number of putatively different subskill function statistically in a very similar manner – it is said that they “load on the same factor” – we have at least to entertain the possibility that they are not different at all, but a single subskill in different guises. If all concievably different subskills load on a single factor, we have to entertain the strong possibility that there are in fact no subskills at all, only a single undifferentiated ability: reading. If some putative subskills function in a statistically similar manner, and so load fairly heavily on one factor, while other putative subskills function statistically in another manner, and so load on a second factor, this is evidence that reading is at least bi-divisible.

For example, Lunzer et al. (1979) is often cited by reading specialists as evidence that it is not possible through test data to differentiate between the so-called sub-skills in reading. This study is said to show that reading (at least as defined by completing reading tests) is a single undifferentiated ability. However, it is interesting to note that while only one principle factor – presumably undifferentiated reading – is identified in this study through factor analysis, there does appear to be some doubt (pp.55-57) concerning the strength of the loading of test items testing word-meaning on that principal factor.

The reader must also remember that Lunzer et al.’s study (as with many of those finding no evidence of multi-divisibility) was conducted on native speakers of English; in fact primary school pupils (presumably still largely free of the specific linguistic problems experienced by some non-native speakers and having crossed the linguistic threshold discussed by Alderson (1984) and Carrell (1991). Our particular concern in this article is different from that of Lunzer et al. in that we are interested in testing adult non-native speakers who will tend to be spread out across the language ability range.

The most recent investigation conducted by Rost (1993), again on native speakers, found strong evidence of unidimensionality, leading Rost to warn against differential skill component interpretation for all available reading comprehension tests (1993:88). However, once again, it is important to note that in the reported factor analysis a second factor which Rost believes to be vocabulary did emerge when the factors were rotated. Rost (1993:80) indeed cites earlier research where “two factors of reading comprehension, namely ‘vocabulary’ or ‘literal reading’ on the one hand, and ‘general reading comprehension’ or ‘inferential reading’ on the other” emerged from the data (Johnson and Reynolds, 1941; Stoker and Kropp, 1960; Vernon, 1962; Pettit and Cockriel, 1974; and Steinert, 1978).

There is further evidence in the literature that the phenomenon of vocabulary loading on a separate factor is not uncommon. Davis (1944) identified two important separate factors in reading as “memory for word meanings” and “reasoning in reading” (a combination of weaving ideas together and drawing inferences from them). Similarly in his later study (1968) a recognition vocabulary test accounted uniquely for a sizeable proportion (32%) of the non-error variance. There is also evidence in Spearitt’s reanalysis (1972) of Davis’s earlier data that vocabulary tests are differentiate from the single basic ability “reasoning in reading” measured by other label reading components in the reading comprehension tests used in the study. Spearitt (1972:110) concluded:

“Vocabulary is the best differentiated, as in both the Davis and Thorndike analyses . . . it could not in fact be subsumed under one general factor with the other three skills”.

Similarly Rosenshine (1980:543) admits to the fact that in three out of the four analyses done on Davis’s data the one unique factor that emerged as separate from the others was vocabulary (“remembering word meanings”), the only exception being Thorndike’s (1973) analysis which he categorises as being less sophisticated than Spearitt’s later study. Rosenshine cites data from Berg (1973) in support of the non-divisibility position but in four out of the five studies summarised by Berg (see Rosenshine 1980:544) lexical competence appears as a separate factor (see also Lennon 1962).

Though the quantitative studies reported above seem to suggest that in general it may not be consistently possible to identify multiple, separate reading skill components, there does seem to be a strong case for considering vocabulary as a component separate from reading comprehension in general. Given that most factor analyses in the studies reported above produced more than one factor, it would be difficult to maintain that reading is a unitary skill. Furthermore, even if the skill components which load more heavily on a second factor also load on the first general reading factor, it might be appropriate only to select test items which load heavily on the first factor when developing a measure of general reading ability. Alternatively,
if vocabulary is considered to be part of reading, a bi-divisible view of reading would seem to be more appropriate.

Qualitative research

As part of a new wave of qualitative investigation in language testing studies, Alderson (1990a; see also Alderson and Lukmani, 1989) investigated the reading skill component question through the judgement of experts on what reading test items actually test. In this study, groups of experts—usually students on MA courses—were presented with a long list of positied reading skill components, and asked to identify cold (“heuristically”) what items in a pilot version of an EAP reading test were measuring in terms of the list. The resulting lack of agreement on assigning particular skills to particular test items, i.e., on agreeing what an item was testing, and even whether an item was testing a “higher level” or “lower level” skill component, could be taken as evidence of the indivisibility of the reading skill, or at the very least could be seen as casting doubt on the feasibility of distinguishing reading skill components. Nevertheless, these conclusions need to be subjected to scrutiny.

The authors of the reading test items used in the Alderson (1990a) study were aware of the possible overlap between skill components tested by individual items. At the time, Weir (1983: 346) had summarised the approach to the design of the reading component in TEEP as follows (for “skills” in this quotation, read “skill components”):

“...we aimed to cover as many of the enabling skills in each of the reading subtests... as was feasible...we indicate opposite each item in the reading sub-tests what the Project Working Party and other experts in the field considered to be the major focus of that item. We were aware that though an item might be seen to be dependant on a particular enabling skill for successful completion, other skills might be contributing to getting the answer right. We realised that the skills we were sampling were not necessarily discrete...”

Thus any conclusions regarding the feasibility of distinguishing separate skill components, based on the inability of judges in the Alderson (1990a) study to agree on what single skill component was tested by individual items, must necessarily be open to question. (For discussion of further weaknesses in this study, see Weir et al., 1990 and Matthews, 1990. See also Lumley, 1993, for a discussion of the centrality of the need for clear definitions and a common understanding of the terms employed, in particular “higher level” and “lower level” skill components, if the attempt to assign skill components to test items is to be meaningful). Furthermore, any similar investigations in this area should ensure that the experts involved share a common understanding of the categories of description employed in the study. There is an alternative literature which suggests that it is possible with clear specification of terms and appropriate methodology for testers to reach closer agreement on what skills are being tested (Anderson et al., 1991; Bachman et al., 1988; Bruten, S.R. et al., 1991; Lumley, 1993; Teasdale, 1989; Weakley, 1993; Weir et al., 1990).

Casting doubt on the unitary nature of reading

It is important to note that, despite the variability in teachers' judgements in Alderson's study, Alderson describes how (1990b:465) “there was more or less agreement” when an item was concerned with deducing the meaning and use of unfamiliar lexical items, or involved understanding relations between parts of a text through cohesion devices. In fact if one were to re-analyse Alderson's (1990b:466) data in terms of our broader categories listed in Table 1 above:

Level (b) Reading carefully to understand main ideas and important detail

Level (c) A knowledge of more specifically linguistic contributory skills

there would be a majority agreement among the judges on which of these levels the items would fall in nine out of the ten cases selected by Alderson (1990:466). If one were to discount the scanning felt necessary to locate one of the items, then there would also be agreement that this item would fall into level (c) type operations. It therefore seems likely that even untrained judges are able to determine when an item is dependent upon specifically linguistic knowledge (level c type operations) as against focusing on more global comprehension (level b type operations). Alderson himself would seem to make this distinction when he classifies without demur gap filling items in the TEEP battery as focusing on lower order (i.e. level (c) type activities (Alderson 1990a:433).

So, perhaps not too surprisingly, it does appear that judges are able to distinguish items which focus upon specifically linguistic knowledge at the word level. Additionally, there is the quantitative evidence from a number of the empirical studies referred to earlier which seems to suggest that these items may well load on a factor separate from that on which other more global items load in reading comprehension tests. Here are two sets of evidence which seem to suggest a bi-divisible view of reading, at least as far as word meanings and reading comprehension in general are concerned.

There is thus both qualitative and quantitative evidence for considering specifically linguistic elements as potentially separable from global comprehension. This is a long way from multi-divisibility, however. There is evidence of a variable use of, and an overlap between, level (a) operations (skimming: reading a text quickly) and level (b) operations (reading carefully to understand main ideas and important detail) used for particular purposes in reading. Different readers would appear to use different strategies/skills to extract the author's message in a text.
(Alderson 1990a:436, 1990b; Storey, 1995; Weir, 1981:34 and 1983:346). There may also be a degree of overlap in some hypothesised skill components (Weir, 1983:346); for example, it is possible that the act of quickly skimming through a passage to get an overall idea of the content will activate various schemata and establish the basis for closer reading by drawing on both the main ideas and important details (see Lee and Musumeci, 1988). Thus the interactive processing model of reading (Buck 1990) talks about in relation to listening may make it difficult in reading to establish the completely separate existence of the constructs included at levels (a) and (b) above.

The ability to perform reading activities on levels (a) and (b) above is obviously dependent on a certain level of competence in the microlinguistic skills detailed as level (c). Alderson (1984:19) termed this "the threshold level" (see also Carrell, 1991; Coady, 1979; Hudson, 1988; Samuels and Kamil, 1984; Stanovich, 1980). It does seem improbable that students would be able to work out the main ideas of a text without some baseline competence in the microlinguistic skills, without understanding some of the relations within at least some sentences of that text (Alderson and Urquhart, 1984; Alderson and Lukmani, 1989; Carrell, 1991; Clarke and Silberstein, 1979; Clarke, 1988; Devine et al., 1987; Eskey, 1988; Grabe, 1991:391; Stanovich, 1980; Storey, 1995; Weir et al., 1990). However, the degree to which the reader needs these lower-order abilities is not yet clear and may prove difficult to quantify (Weir et al., 1990:508). An interactive or interactive-compensatory view of reading would seem to imply that readers can make differential use of a range of skill components which we might loosely label "specifically language related" and "reason related" (Grabe, 1991; Williams and Moran, 1989). Given the evidence against both fully unitary views of reading ability on the one hand, and multi-divisible views on the other, it is important for those concerned with language test development to reflect critically on the ramifications of operationalising either in reading tests.

**DANGEROUS IMPLICATIONS**

Dangerous implications of a fully unitary view

"There is a disturbing corollary of the fully unitary argument which deserves serious attention from all involved in developing language tests. If there are no discernible skill components in reading — it is unidimensional — it should not really matter how we test it, or what operations we try to assess. The inability to provide consistent or conclusive empirical evidence (either quantitative or qualitative) for the separability of skill components would make it difficult to utilise test formats with a specifically linguistic focus (level (c) type operations in Table 1 above). These are often relatively easy to construct, administer and mark, frequently have respectable psychometric properties and reliability estimates (usually high internal consistency estimates), and frequently correlate fairly highly with more global tests of reading comprehension.

Given that the current consensus on the nature of reading includes microlinguistic elements as important components within reading (Grabe, 1991, and Williams and Moran, 1989), and given the evidence for a unitary view of reading (Lunzer et al., 1979, Rosenshine, 1980; Rost, 1993), it is perhaps not surprising that many current reading tests have in fact ventured down this microlinguistic road (quite often for reasons of practical expediency rather than from a principled view of unidimensionality). In the British Council/UCLES IELTS test a third of the items in some of the reading modules appeared to be testing at the microlinguistic level, and in the G1 General Reading Comprehension Module, Criper and Davies (1988:89-97) indicated that over 50% of the items focused on such lower order elements. The more recent IELTS test has a number of items at this level (see Weir, 1990); so do the TEEP test (see Weir, 1990), TOEFL, and many other tests of international standing and good repute. It is thus fairly urgent that the status of such lower order elements in the measurement of reading ability be investigated.

There are serious question marks against the value of testing directly the specifically linguistic level (c) operations) listed in Table 1 above. The evidence from the literature reported above and our own initial investigations throw some doubt on the value of including any items which focus on specific linguistic elements (e.g. individual words or cohesive devices) in tests which purport to make direct statements about a candidate’s reading ability. Tests focusing on level (c) microlinguistic elements may well correlate quite highly with test which attempt to tap into global facets of that ability, but they do appear to load on a different factor from general reading comprehension in many studies. It is in our opinion unarguable that level (c) elements contribute to level (a) and (b) abilities, but on their own they may not constitute an adequate predictor of that ability.

The limited data reported below (pp 10-14) are intended to raise awareness of the possible danger that some candidates might be seriously disadvantaged by the inclusion of such discrete linguistic items in tests of reading comprehension (level (c) operations in Table 1 above) where the purpose is to indicate whether a candidate has understood the main ideas and important detail provided by the writer, i.e. the writer has understood the text levels (a) and (b) operations in Table 1. Further data sets are also cited which cast doubt on the use of specifically linguistic test items and the formats these are commonly associated with to test reading ability.

We are aware that our argument smacks of the assertive at present and that further data will need to be generated to prove or falsify our fears. It is a purpose of this paper to generate interest in this issue so that a concerted body of empirical research might throw further light on it.
Dangerous implications of a multi-divisibility view

The dangers of a unitary view of reading have been outlined above, in the light of evidence concerning the status of items testing at the specifically linguistic level. Equally serious problems may arise in accepting unthinkingly a multi-divisibility view. This "scattergun" approach to the testing of skill components in reading, as referred to above in connection with the original TEEP test (Weir, 1983; 1991) likewise needs to come under close scrutiny. In the past, broadly sampling skill components across the three levels (a), (b) and (c) seemed a sensible course of action for assessing reading comprehension. By attempting to sample across the range of hypothesized "subskills", the intention was to take an adequate sample of a construct that could be labelled reading (Weir, 1983; 1990).

The University of Reading data

Doubts now arise concerning the relative contribution of skill components to measure the construct of reading, at least in terms of the relationship between level (c) and the other two levels. In the placement assessment (broadly styled on the TEEP approach described above) administered to students recently entering the presessional EAP courses at Reading University in July and August, (candidates arriving range from roughly around band 5 on the ELTS/EILTS scale to band 6.5), it was noted that there were a number of students who might be able to cope quite well with reading passages and questions at the global level, but that this was not matched by their performance on test items focusing on more microlinguistic items: cohesion markers, discourse markers, lexis and structural elements. Figure 1 below plots the relative performance of students on global short-answer and microlinguistic gap-filling item-types respectively. It will be seen that students scored overwhelmingly in the top 50% of the available score range on global items; scores on microlinguistic items, on the other hand, were more evenly distributed on either side of the 50% line, with a slight tendency to fall below rather than above that line. It will also be noted that, whatever the item type, very few students achieve more than 80% of the available score range for microlinguistic items, whereas in each case substantial numbers of students achieve more than 80% of the available score range on global items. This situation initially gave pause for thought as to how to allocate such students to classes on the basis of their performance on/reading test items.

However, of far greater concern were the implications of using specifically linguistic test items later on in the programme for assessing proficiency in reading prior to entrance into the university proper. Here there was a serious dilemma in interpreting test results. Doubts arose about the wisdom of using tests or items which appeared to focus on microlinguistic elements. To be more precise, there was major concern about the fairness of such tests for the subset of students who did "well" on the global comprehension items in our reading tests (at the moment for descriptive

Figure 1: Student performance on global as against more specifically linguistic gap-filling items (15 global items, 39 microlinguistic items)
purposes arbitrarily defined as more than half the items right) but did not do so well on the microlinguistic items (less than half the items right). Such a situation may occur for a variety of reasons, e.g. successful application of background knowledge to the text under review and/or transfer of higher level processing skills from the L1 which compensate for deficiencies in lower level linguistic abilities (Coady, 1979; Goodman et al., 1979; Hudson, 1988). This group of candidates must in any case have passed a threshold level of language proficiency necessary for this transfer to take place (Alderson, 1984; Clarke, 1988; Cziko, 1980; Devine, 1987; 1988) in order for global questions to be correctly answered.

The low loading of vocabulary on a general reading comprehension first factor in some of the analyses reported earlier in this article is consistent with these findings, in that such results may reflect that candidates can comprehend the overall ideas in a passage irrespective of specific linguistic shortcomings. Additionally, where these specific items are known already by a candidate the ability to answer questions set on them may have very little to do with understanding the passage they occur in.

The evidence presented in Figure 1 will need to be supplemented by further, more detailed studies. What an adequate performance on the two item types is will need to be more precisely operationally defined. Simply taking a 50% cut-off point on each test, whilst useful for descriptive purposes at the early stage of our enquiry, is rather crude. Furthermore, it is possible that a group of candidates might randomly appear in the top left-hand quadrant as a normal feature of score distribution. Moreover it needs to be investigated whether the same candidates are similarly distributed over a number of measures each of the bi-divisible elements identified.

There is also a possibility that the data we present above may have been unduly influenced by the formats employed, i.e. those in the top left hand quadrant in Figure 1 may have reacted adversely to the gap-filling format used to assess microlinguistic knowledge. However, synthesis of information available from a variety of sources would seem to suggest that format effect would be unlikely to account for our results. Evidence from short-answer and multiple-choice formats indicates that the items in reading comprehension tests which focus on microlinguistic elements such as lexis or cohesion do not necessarily contribute to the overall measurement of reading in ways similar to items which test more global comprehension, in that they can be shown to correlate more highly with their own subtest or with other microlinguistic elements in a test battery (such as separate subtests of vocabulary and structure).

Data from ESP Centre, Alexandria, Egypt

Similar differential performance is emerging in data from a battery of EAP tests under development by the Testing and Evaluation Unit at the ESP Centre in Alexandria, Egypt. (An example of these data is provided in Figure 2 below.) These are short answer questions on the same passage designed to test separately reading comprehension at levels (a), (b) and (c) in Table 1 above. Point biserial correlations show that items in section 1 of the battery testing at levels (a) and (b) correlate more with their own subtest than they do with level (c) (the microlinguistic items), and vice versa. Similarly, there is a small subset of students who, whilst coping well with (a) and (b) operations, experience more difficulty with the specifically microlinguistic items. So here again there is the same phenomenon of differential performance on global as against specifically microlinguistic items. It should be noted that in the Alexandrian case, the reading test format differed from that used in the Reading study. (An example of these data is provided in Figure 2).
Data from College English Test (CET) China

Evidence of differential performance appears in further test contexts. Items which focus on cohesion or working out the meaning of words in context appear to be out of place in a recent version of a reading component of the College English Test (CET) used to test hundreds of thousands of undergraduates across China. Such items according to item/part correlational data (Pearson Product Moment) would appear to relate more closely to a later section in the test which focuses on vocabulary and structure than they do to the section designed to test reading comprehension. The format used in this test is multiple-choice, so the phenomenon would appear to repeat itself here in a third format. This is further evidence that the results obtained on the microlinguistic as against global focus are probably not the effect of item type.

Data from Alderson and Lukmani (1989)

Alderson and Lukmani (1989: 269) found that:

"...what seems to be happening is that weaker students overall do somewhat better on higher order questions than lower order questions... perhaps lower order questions measure language ability whereas higher order questions might be said to measure something like cognitive skills, logic, reasoning ability and so on...it might perhaps be possible to get a better estimate of a student's reading abilities... from higher order questions rather than from lower order questions... one should not perhaps believe that students with lower language levels are incapable of answering higher order questions. In other words, one should not be inferring from poor performance on lower order questions an inability to perform well on higher order questions".

CONCLUSIONS

The argument as to whether reading is multi-divisible, consisting of a number of skill components which can be identified clearly, or whether it is an indivisible, unitary process, is still unresolved. If a fully unitary view is to be clearly rejected, future research will need to demonstrate the consistent presence of at least a second skill component in repeated analyses across a range of samples of ESOL candidates. Secondly, future research will need to investigate whether such components are identifiable. Finally it will have to establish the extent to which each component has a meaningful effect on the measurement of reading comprehension. How much of the overall variance does each component explain in a reading test? It will be important to use more exigent statistical techniques to test the presence of each component is statistically significant.

There is cause for immediate concern that wholesale adherence to either the unitary or the multi-divisible view in language testing may be problematic. For further investigation of these matters, it will be necessary to develop tests which are maximally valid tests of the skill components at levels (a), (b) and (c) (through systematic a priori validation including expert judgement and student introspection) and to investigate the performance on these tests of students at a variety of ability levels.

It may of course prove impossible to operationalise skill components at the three levels separately in a test. It may be that reading is such a massively parallel interactive process that we will not be able to distinguish clearly between components. It may be that at certain levels of ability, for example weak and strong readers, reading is indeed unitary; divisibility may be a function of the level of student being tested. For readers linguistically proficient in the target language and already competent readers in their L1, reading in the target language may well be unicomponential, whereas this may not be the case where either of these conditions is not met (see Downing and Leong, 1982).

It does appear, however, that there is doubt about the status of items which focus on specifically linguistic operations at level (c) as part of the assessment of a candidate's general reading ability. As a matter of urgency, it is necessary to investigate whether testing at level (c) does in fact give us sufficient information about a candidate's ability to handle level (a) and (b) activities. We must address the implications of emerging evidence that there may be groups of candidates who are capable of operating at levels (a) and (b) but who are severely disadvantaged by test items which focus on level (c). There is a suspicion that test items which focus on the specifically linguistic/individual word level may not be good predictors of general reading ability, i.e. they do not give us an accurate picture of the reading ability of all the individuals who sit a test.

In proficiency tests, the issue of validity is crucial given the use to which the results of such tests are normally put. If items with a specific linguistic focus are used in the measurement of reading ability, and if the results are then used in taking decisions on entrance to further study or to the professions, unfairness might result. Tests including such items might discriminate against the linguistically disadvantaged but otherwise competent reader. Those students who can understand almost all the main ideas and important information in a passage but who are unsure about the meaning of particular lexical items or cohesion devices - items selected for inclusion perhaps subjectively or idiosyncratically - may not pass through the entrance gate which the test embodies. To this extent the idea of reading comprehension as either purely unitary or multi-divisible would appear to be dangerous.

BIBLIOGRAPHY


Stanovich, K.E. (1980). "Towards an interactive compensatory model of individual differences in the development of reading fluency" *Reading Research Quarterly* 16/1:32-71


Teasdale, A. (1989). "Introspection and Judgemental Approaches to Content Validation: A Study Using the Test in English for Educational Purposes" Unpublished Dissertation MAAL, Department of Linguistics University of Reading


Weakley, S. (1993). "Procedures in the Content Validation of an EAP Proficiency Test of Reading Comprehension" Unpublished Dissertation MATEFL CALS University of Reading


