

Reading Skills: Hierarchies, Implicational Relationships and Identifiability.

C.J. Weir, A. Hughes and D. Porter

University of Reading

This paper is a direct response to the first part of Alderson's paper, published in *Reading in a Foreign Language* 6.2. Weir et al criticise Alderson on two main points: (a) they claim that his methodology was defective, in that no evidence is given that his judges had any experience in the task they had to do, they were given no training, and the terms 'High' and 'Low' order questions were insufficiently defined; (b) Weir et al criticise Alderson's view of the implicational relationships between 'High' and 'Low' order questions, and the assumptions he makes about their relative difficulty. They argue that, while the validity of the 'High/Low' distinction is an open question, it is perfectly possible for practical testing purposes to obtain reliable judgements from properly selected and trained judges.

INTRODUCTION

In an article in the previous issue of this journal, Alderson (1990) reported on investigations that he had carried out into a number of issues involved in the testing of reading comprehension skills. As the issues he addressed are undeniably important and worthy of serious investigation, and as Alderson's findings lead him to draw conclusions which have fundamental implications for the construction of valid reading comprehension tests, and perhaps also of tests of other linguistic skills, the article calls for careful critical attention. Unfortunately, it is our opinion that such attention appears to reveal mistaken thinking at a number of crucial points in the article, weaknesses in the methodology, and a number of inaccuracies. The cumulative effect of these is to undermine Alderson's conclusions.

Alderson's article is concerned with the identifiability of separate reading skills, the categorization of these into "higher order" and "lower order" skills, and the question of whether implicational relationships may be said to hold between these two categories. His preliminary conclusions are that:

- i. Judges are unable to agree as to what an item is testing
- ii. Judges are unable to agree upon the assigning of a particular skill to a particular test item.
- iii. Judges are unable to agree upon the level of a particular skill or a particular item.
- iv. There appears to be a lack of relationship between item statistics and what an item is claimed to be testing.

and the following explanations are offered:

- a. There are serious reasons for doubting whether a skill can be said to be "higher" or "lower" than another skill in any hierarchy that implies relative difficulty or some differential stage of acquisition (at least for ESL readers).
- b. The skills identified by Weir (1983) and Criper and Davies (1988), both based on Munby (1978), are in fact overlapping rather than discrete.
- c. It is inappropriate to categorise the skills in the Munby taxonomy into "higher" and "lower" order skills.
- d. It is unlikely that any test item can be unambiguously said to be testing any one skill.

SKILL HIERARCHIES: "HIGHER ORDER" AND "LOWER ORDER" SKILLS

As Alderson notes, it is indeed common practice among teachers, testers and researchers in the field of reading to consider reading skills as being arranged hierarchically into "higher order" and "lower order" skills. So long as some such distinction is valid, and whatever the actual relationship between the two orders of skill, it will be necessary for a valid test of reading comprehension to be based on an appropriate sampling of these higher and lower order skills (although in some interpretations "appropriate sampling" might be taken to imply a neglect of lower order skills, these being subsumed in the higher order skills measured). It would then be reasonable to take the notion of appropriate sampling as implying that it should be possible to say *a)* what the relevant skills are, *b)* which ones are higher and which lower order, and finally *c)* which of these skills or which combination of them is being tested by a particular test item. Alderson seeks to shed some light on the second and third of these questions.

Alderson's basic and disturbing finding is that judges do not agree on whether a given skill is higher or lower order, or on what skill is being tested by a given item. This leads him to speculate on the legitimacy of the higher/lower order distinction, and of the so-called reading enabling skills proposed by Weir (1983), Criper and Davies (1988) and Munby (1978). It seems to us, however, that the nature of the judgement exercise whose results lead to this speculation, as described in Alderson (1990), may be fundamentally flawed on three main grounds, and thus that the speculation is at least questionable on the evidence provided.

Firstly, we would question the nature of the judges, said to be “18 experienced teachers of ESL”. As language testers are well aware, experience does not guarantee reliability of judgement (particularly when the experience was not gained in making judgements!), but no evidence is adduced in this investigation for the reliability of the teachers in making the type of judgement required. Without evidence that they were indeed reliable judges, we should perhaps not be surprised that the teachers did not agree on the various types of judgement they were called upon to make in this case. No more should we be surprised that they changed their minds on judgements previously made. If confirmation of the possibility of consistent judgements were sought, the best possible judges should have been employed – judges with relevant experience and for whom some evidence of appropriate reliability could be shown. Future investigations of this type should thus employ judges with some track-record of demonstrated reliability; suitable judges might be teachers or testers or even experienced general linguists, but their qualification for the task should not be in doubt.

Secondly, it appears that the teachers in the investigation were being asked to designate enabling skills as higher-order or lower-order, when no firm definition of these terms was given. We are told that teachers were asked to make their judgements “after discussion of the posited distinction between higher and lower order skills, and its *possible* nature...” (our italics). While the ambiguity of the phrasing leaves it unclear whether the teachers were invited to participate in the discussion and thus in the speculation as to the nature of the distinction they were being asked to make, there is evidently no certainty in the minds of either the author or the teachers as to what that nature is. Once again, it is not surprising that there should be evidence that the teachers changed their minds over what was higher- or lower-order. The impression of uncertainty is reinforced by Alderson’s reference to Alderson and Lukmani (1989) – a fore-runner of Alderson (1990) – who, in discussing the results of an investigation similar to the one currently under discussion, “speculate that lower order questions might be said to measure language abilities whereas higher order items might measure cognitive skills, reasoning ability, and the like”. They thus wonder after the investigation what it was that they were investigating! The use of undefined key terms in this way places judges in an Alice-in-Wonderland-like world, and it is no wonder that there should be little consistency of judgement among them or that, where there is consistency, it should produce such odd decisions as that “Understanding explicitly stated ideas” is a lower-order skill. (Indeed, had the teachers been directed to the work of Alderson (1978: 366, 372–373, etc.), where “lower-order” is used to refer to specifically linguistic relations at sentence-level and below – syntax and lexis – while “higher order” is concerned with relations above sentence level – making inferences, handling text, working out the main idea – one would have expected a much greater consistency

in the identification of skill-levels). Any future investigation in this area should begin with a solid and serious attempt at achieving a maximally clear definition of the terms "higher level" and "lower level".

Finally, it is axiomatic that, if consistency of judgement is to be achieved, potential judges will need training for the task they are to perform. Alderson (1990) speaks only of a "discussion of the posited distinction between higher and lower order skills, and its possible nature". This cannot be regarded as adequate training, either in its nature – or in its likely duration. Once again, it is not surprising that the use of teachers untrained in making judgements such as these is associated with lack of consistency in the judgements obtained, both from teacher to teacher, and from occasion to occasion in individual teachers.

Thus Alderson's third conclusion that "judges are unable to agree upon the level of a particular skill or a particular item" has to be evaluated in the light of the above weaknesses in methodology, and cannot be said to have any clear implications for the nature of the enabling skills or of the higher/lower order distinction.

IMPLICATIONAL RELATIONSHIPS BETWEEN SKILLS

For the purposes of the present discussion let us accept as an established definition Alderson and Lukmani's "speculation" that lower order skills are those which deal with language abilities, whereas higher order skills are concerned with cognition, reasoning ability, etc. According to Alderson, not only is the view widely held among reading specialists that reading skills may be arranged hierarchically into "lower order" and "higher order" skills (e.g. Alderson, 1978), but also that lower order skills are necessary before higher order ones can be acquired or developed. It is not clear to us, however, how widely this latter view is held by language testers.

What we believe that language testers would more readily admit to is the idea that, in the processing of text, the exercise of higher order skills will naturally involve the use of some lower order skills, regardless of order of acquisition. For example one cannot expect to obtain the gist of a text, or understand the main ideas (higher order skills) without understanding some of the relations within at least some sentences of that text (lower order skill). The acquisition of particular higher-order skills in the foreign language, on the other hand, may not be entirely dependent on the foreign language itself, as they may be inherited directly from use with the mother tongue.

But which of the two above views is held (the acquisitional sequence view, or the use-involvement view) does not affect the argument presented here, since neither of them is addressed by Alderson's investigation. He attempts rather to approach the issue of implicational relationships between skills by examining patterns of response to items in the reading sections of two English language proficiency tests (TEEP and ELTS). Alderson appears to believe that if there is an implicational relationship

between high- and low-order skills of the kind referred to above, then test items that aim to measure higher-order skills should be more difficult than those that aim to measure lower-order skills. In other words, higher order skill items should have lower facility values than lower order skill items.

It is not clear why this view should be held. Except in very exceptional circumstances, unlikely to be met in an operational test, *there is absolutely no reason to expect the higher or lower order of skills tested by two items to determine the relative difficulty of those items.* To give a simple example, failure to decode the word “excaudate” in this sentence (lower order skill) does not imply failure to obtain the gist of this paper (higher order skill). An item testing the first of these could well have a lower facility value than one testing the latter. But no-one would conclude from this that the decoding of words is unnecessary in order to obtain the gist of this paper. The question of implicational relationships between skills is thus not properly addressed through an analysis of facility values.

In a similar manner, Alderson appears to expect that if there are implicational relationships between skills, then lower order items will consistently discriminate better than higher order items. We can find no logical basis for such an expectation, however. Discrimination indices will depend crucially on the specific nature of the individual items, and to some degree on their level of difficulty in relation to the ability range of the sample of subjects taking the test. In short, Alderson’s fourth conclusion that “there appears to be a lack of relationship between item statistics and what an item is claimed to be testing” is unsurprising, and fails to throw any light at all on the question his research is meant to address.

THE IDENTIFIABILITY OF SKILLS

We said above that we thought most testers of reading comprehension might agree that the exercise of higher order skills might involve the use of lower order skills. It follows from this that an item testing a higher order skill must inevitably, indirectly, test lower order skills. Indeed, all but the lowest order of items (for, as we saw above, there is no sound reason for assigning skills to two order categories only) will involve the simultaneous testing of lower order skills. It is not surprising, therefore, that there was lack of consistency among the teachers in Alderson’s study on what single enabling skill and what order of skill an item was testing, when the item was testing more than one skill, both higher and lower order. The language tester’s – and test consumer’s – primary interest will frequently be in the highest order of skill being tested. Particularly where this is the case and is made clear, where we can have confidence in the reliability of the judges, where clear examples are provided of items testing specified skills, and where judges participate in training for standardization, very high levels of agreement between judges can be achieved

(Weir, Porter, Hughes and Laboso, forthcoming). Where the test constructor's focus is on lower order, more strictly linguistic skills, high levels of agreement can be achieved if the above conditions are fulfilled.

CONCLUSION

The questions of the legitimacy of a higher/lower order distinction in reading skills, of the relationship between these orders if they can be substantiated, and of the identifiability of individual skills, clearly remain important both for a better understanding of the nature of reading and for the construction of valid tests of reading comprehension. The research reported in Alderson (1990) makes little contribution to the resolution of these questions.

BIBLIOGRAPHY

- Alderson, J. Charles (1978) *A Study of the Cloze Procedure with Native and Non-native Speakers of English* Unpublished Ph.D. Thesis, University of Edinburgh.
- Alderson, J. Charles (1990) "Testing Reading Comprehension Skills (Part One)" *Reading in a Foreign Language*, Vol.6 No.2 pp425-438.
- Alderson, J. Charles and Yasmeen Lukmani (1989) "Cognition and Reading: Cognitive Levels as Embodied in Test Questions" *Reading in a Foreign Language*, Vol.5 no.2 pp253-270.
- Criper, C. and A. Davies (eds) (1988) *ELTS Validation Project Report*. English Language Testing Service Report Col.1(i). The British Council and the University of Cambridge Local Examinations Syndicate.
- Munby, J.L. (1978) *Communicative Syllabus Design* Cambridge University Press.
- Weir, C.J. (1983) *Identifying the Language Problems of Overseas Students in Tertiary Education in the UK*. Unpublished Ph.D. thesis, Institute of Education, University of London.
- Weir, C., Porter, D., Hughes, A. and Laboso, J. (forthcoming): "Achieving consistency in test-content analysis."