Skill Taxonomies and Problems for the Testing of Reading

Margaret Matthews
Centre for Applied Language Studies, (CALS) University of Reading

This is another response to Alderson's paper. Like Weir et al, Matthews is critical of aspects of Alderson's methodology. She however devotes more attention to what she considers the inadequacy of taxonomies of skills to describe individual readers' processes, and hence their usefulness in test construction.

INTRODUCTION

The recently reported studies by Alderson and Lukmani (1989), and Alderson (1990), of judgements concerning the relationship between micro skills and test items raise a number of issues. Some of these relate to the studies in question and others to the nature of micro skills and reading tests generally.

As the second study was in some respects intended as a refinement of the first, I will focus on particular aspects of the 1990 paper, namely the hypotheses which were examined, the conclusions which were drawn, and the research method itself. I will then go on to discuss skill taxonomies and reading processes generally, and finally to draw tentative conclusions for proficiency testing.

ALDERSON'S PROCEDURES

Alderson gave a group of informants a taxonomy of micro skills and asked them first to judge the level of each one (higher or lower), then to match a number of test items to each skill, and finally to ascribe levels to the test items. He reports a considerable lack of consensus concerning the various tasks, and draws various conclusions from this. However, although the conclusions may be valid inferences, it seems likely that various aspects of the procedure itself would have contributed to the lack of consensus.

Firstly, Alderson refers in his introduction to a *scale* of skills, but asks his judges to make *binary* distinctions. If the skills are, in fact, relative, then the task of assigning them artificially to two absolute categories would surely contribute to a level of disagreement.

A further constraint upon task performance would be a lack of clear agreement concerning the definition of higher– and lower–order (skill). Although there was, it seems, some preliminary discussion about the "possible nature" of the distinction,

Margaret MATTHEWS teaches EFL at the Centre for Applied Language Studies, University of Reading.

there is no statement in the report of the study about which, if any, working definition was arrived at. This would be crucial for the reliability of results as informants may have been orienting to different characteristics such as degree of difficulty, degree of value or entailment. That the definition may have been somewhat fluid even for individual informants is clear from the fact that, as Alderson points out, matches which were made between skill descriptions and level were not always consistent with matches, made by the same person, between skills underlying particular items and level.

CONCLUSIONS AND HYPOTHESES

The second part of Alderson's study compares the more consistent judgements about the level of test items with the actual difficulty of these for testees. No clear relationship is found between the two and Alderson regards this as evidence for doubting that the micro skills are differentially acquired. This seems to be a fair interpretation of the results as far as the group of testees as a whole and the posited level of skills is concerned. However, although it was clearly outside the scope of Alderson's study, an examination of the responses of individual testees might reveal interesting developmental patterns.

One debatable point concerns the hypotheses, which Alderson seeks to disprove, that if taxonomies involved an implicational scale, the items deemed to be lower-order would prove easier than items deemed to be higher-order, and his conclusion (tentative) that no evidence for such a scale is provided. I would like to cast doubts on both the hypothesis and the conclusion.

Firstly I would expect that if an implicational scale is involved (and I believe it is) then so called lower-order items would probably be **more** difficult than so called higher-order items, for reasons which I will elaborate below. Secondly, disregarding the unreliability of the judgements to which Alderson draws attention earlier, the conclusion could be that there is more to reading than this list of micro-skills, and that other factors contribute to the comprehension of test items.

In his introduction Alderson remarks that a hierarchy of skills is commonly implied by reading specialists, who refer to "higher" and "lower" skills. He goes on to caution that "an implicational scale is not a necessary consequence of a hierarchy" (opp cit p425), and gives the example that the "ability to evaluate" may not depend upon the prior possession of an "ability to understand explicitly stated information". But, whilst the former, general observation is presumably correct, the latter example surely is not: *in relation to any particular text* evaluation *does* logically entail understanding. Indeed this and other implicational relationships can be discerned in the various taxonomies, deriving from Munby (1978), which proliferate: most obviously, for example, "understanding implied meaning" presupposes an "understanding of explicit meaning" and being able to "distinguish main ideas"

crucially depends on several other listed "skills" such as "understanding relations between parts of a text". The serious problem for measurement which such interrelationships cause will be referred to later.

THE NATURE OF MICRO - SKILLS

Although Munby's general approach to course design has not escaped criticism, (see e.g. Widdowson 1980; Holliday and Cooke 1982; Hutchinson and Waters 1987) taxonomies deriving fairly directly from his are now commonplace in teaching and testing, and have given rise to little debate concerning their suitability as a basis for test content specifications. Indeed a version of these has been retained by the new IELTS test. They do, however, warrant closer scrutiny.

Munby's categories relate, with few exceptions, to levels of linguistic organisation, and are arguably knowledge– rather than skills–based. His list represents a clear sequence from smaller to larger linguistic units and as such contains strings of implicationally related categories. It is for this reason that, where tests are based upon such a taxonomy, I believe "higher-order" items will be easier, *other things being equal,* than "lower-order" items, because with any text of a reasonable length there is so much redundant information available for performing 'global' tasks. What one might have lost on a roundabout one could compensate for on a swing. This point about difficulty levels will be returned to later.

Although Munby refers to his categories as "skills" the precise status of these, as mentioned above, is unclear. They are expressed in terms of an operational element and a communicative feature (opp cit p117), and in the case of the receptively formulated categories one wonders whether the operational slot could be realised by the verb "Knowing". Thus "Knowing the meaning of the word 'tree' " might be a candidate for inclusion. Such a description seems to differ only in terms of its specificity from "Understanding explicitly stated ideas". Any given reader presumably understands some ideas and not others, and no amount of practice will change this situation.

If most of the categories, as it appears, relate more to knowledge than to skill, the taxonomy might be a suitable source for the specification of structural, discrete-point tests. But such tests have been largely abandoned in recognition of the fact that there is more to reading than a knowledge of structure, essential though this might be.

OTHER FACTORS UNDERLYING THE READING PROCESS

It is now a commonplace that categories of knowledge other than linguistic ones are essential for successful comprehension: the development of schema theory has demonstrated the importance of different sorts of background knowledge. But in addition to *knowledge*, what sort of *skills* are required? Eskey and Grabe (1988) emphasise the importance of speed and automaticity. For example, in the case of

vocabulary they claim that it is necessary but not sufficient for successful reading to *know* words: the skill of instant recognition must also be developed. They and others (Carrell 1988; Devine 1988) also emphasise the importance for success of coordinating "top down" and "bottom up" processing, and the negative consequencies of developing the one mode at the expense of the other. Thus although the reading process is little understood, suggestions such as these, found in the reading literature, intimate the likely complexity of the abilities involved.

SKIMMING AND SCANNING

Coverage is invariably given in the taxonomies to "skimming" and "scanning", but these seem to be rather different in kind to the other "skills". Are they in fact skills, or are they strategies, planned approaches to a given task? (see e.g. Thomas and Augstein 1984) It is difficult to see how skill is involved here except in terms of flexibility and speed: presumably even pre-readers can scan successfully given unlimited time. And if speed is an important criterion then presumably ability in these two areas must be tapped separately by sub-tests in which speededness is an important element.

At any rate the items in Munby-based taxonomies appear to be a slightly random and overlapping collection of strategies, skills and (chiefly) knowledge, and represent an impoverished account of the reading process.

DIFFERENTIATED PROCESSES

One further issue, mentioned by Alderson, arises out of his study. It is hardly surprising that judges did not fully agree upon the skill which specific test items were tapping, for apart from anything else there are many roads to Rome, and who can say exactly in the case of any individual (whether test constructor, judge or testee) which particular route might be taken. If this point is unconvincing, a consideration of the popularly included micro-skill "Deducing the meaning of unfamiliar lexical items" should be persuasive. It is clear that individual testees have different lexicons, and that in the case of any given item guesswork may not be necessary for some, while for others different parts of the context may provide the necessary clues. The test constructor cannot possibly predict, unless language has been learned in laboratory conditions, whether that particular skill will be crucial for a testee. Furthermore (pace Alderson opp cit p343) it is doubtful whether judges or testees can retrace the particular route taken by themselves with any degree of reliability. Accurate self-analysis is always difficult, and statements which involve bringing to consciousness processes which normally operate below this level must, I think, be regarded with great caution.

DIFFERENTIATED PRODUCTS

The argument about variability in reading processes within and between individuals has been taken one step further by many who claim that the product itself is variable. The view that meaning does not reside in text, but is created by individual interaction with the text, has long been orthodoxy. As Widdowson (1979) expresses it, "since conceptual worlds do not coincide there can never be an exact congruence of coder's and encoder's meaning." (p180). It follows that there cannot be an exact congruence of test constructor's and testee's meaning. Nor does this lack of congruence necessarily occur at subtle, abstract levels only. Urquhart (1987) is surely not alone when he reports consciously wondering which possible response test-setters might require.

CONCLUSION

I believe that certain aspects of the procedure adopted by Alderson would have contributed substantially to his findings. I also believe that the assertions, assumptions and conclusions contained in his report are, in part, mistaken. Nevertheless I share the distrust of "skill" taxonomies which presumably motivated the studies in question, and propose the following considerations for reading test design:

1. It is likely that the knowledge and skills which underlie reading processes are more numerous and complex than currently prevalent taxonomies allow, and furthermore that various kinds of knowledge and skill interrelate differently in the case of individual readers, texts, purposes etc.

   Teachers will continue to focus on any aspect of reading which they judge to be necessary, and achievement tests will reflect the teaching accordingly. But the construct validity of proficiency tests based upon a selection of items from Munby must be questionable, and for those tests which have a paramount placement or screening function this lack of validity is clearly unacceptable. Given the present limited state of knowledge it might be prudent to set current enthusiasm for processes aside, and test products. As Hughes (1986) points out, if texts are carefully chosen and tasks are sufficiently global then the relevant enabling skills will be naturally sampled.

2. The difficulty concerning variability of product is hard to address, but the problem can be minimised by trialling tests thoroughly with native readers, to identify items which are most likely to give rise to incongruent responses.

3. All other things being equal, items relating to large stretches of text ('global' items) are likely to be easier than items relating to small stretches of text. However, other factors such as the application of culturally inappropriate

schemata (Carrell and Eisterhold; 1988) or a lack of content knowledge (Alderson and Urquhart 1983) may cause readers to make false assumptions and respond incorrectly.

4. Where test specifications involve one or more implicational scales and items are scored incrementally, marking is inaccurate, as the same abilities are credited more than once.

BIBLIOGRAPHY

Alderson, J. Charles (1990) Testing Reading Comprehension Skills *Reading in a Foreign Language* Vol.6 no.2 pp425–438.

Alderson, J. Charles and Lukmani (1989) Cognition and Reading: Cognitive Levels as Embodied in Test Questions. *Reading in a Foreign Language* Vol.5 no.2 pp253–270.

Alderson J.C. and A.H. Urquhart (1983) The Effect of Background Discipline on Comprehension – A Pilot Study. In A. Hughes and D. Porter (Eds) *Current Developments in Language Testing.* London Academic Press.

Carrell, P. (1988) Interactive Text Processing: Implications for ESL/Second Language Reading Classrooms. In P. Carrell. J. Devine and D. Erskey (Eds) *Interactive Approaches to Second Language Reading.* New York: CUP.

Carrell, P. and J. Eisterhold (1988) Schema Theory and ESL Reading Pedagogy. In P. Carrell, J. Devine and D. Eskey (Eds).

Devine, J. (1988) The Relationship between General Language Competence and Second Language Reading Proficiency: Implications for Teaching. In P. Carrell, J. Devine and D. Eskey (Eds).

Eskey, D. and W. Grabe (1988) Interactive Models for Second Language Reading: Perspectives on Instruction. In P. Carrell, J. Devine and D. Eskey (Eds).

Halliday, A. and T. Cooke An Ecological Approach to ESP. In *Issues in ESP: Lancaster Practical Papers in English Language Education:* A. Waters (Ed) 1982 Pergamon.

Augstein, S. and L.F. Thomas (1984) Conversational Investigations of Reading: the Self Organised Learner and the Text. In J.C. Alderson and A. Urquhart (Eds) *Reading in a Foreign Language.* New York: Longman.

Hutchinson, T. and A. Waters (1987) *English for Specific Purposes.* CUP p54.

Hughes, A. (1986) A Pragmatic Approach to Criterion – Referenced Foreign Language Testing. In M. Portal (Ed) *Innovations in Language Testing.* Exeter: NFER Nelson.

Munby, J. (1978) *Communicative Syllabus Design.* Cambridge: CUP.

Urquhart, A. (1987) Comprehensions and Interpretations. *Reading in a Foreign Language* Vol.3 no.2 pp387–410.

Widdowson, H.G. (1980) ESP: The Curse of Caliban. *TESOL.* San Francisco.

Widdowson, H.G. (1979) The Process and Purpose of Reading. *Explorations in Applied Linguistics.* London: OUP.