

Testing Reading Comprehension Skills (Part Two) Getting Students to Talk About Taking a Reading Test (A Pilot Study)

J. Charles Alderson
University of Lancaster

In the second half of his paper, Alderson reports on introspective and retrospective accounts from test takers, in a further attempt to examine the validity of reading, and test-taking "skills".

Since a body of data had been built up on the TEEP test T012, it was decided to use items from this test in the pilot study. The test consists of a series of short-answer questions based upon a seven-page text entitled "Smoking and Other Addictions". The text is taken from an academic journal article (Weir 1983). The original test contained 15 items, which was felt to be too long for an introspective study. Ten items were therefore selected from the fifteen, and were retyped into a test booklet. In what follows, the item numbers refer to the pilot test, not the original TEEP test. (The original TEEP item number is given in parentheses in Table 8).

In constructing this pilot test, an attempt was made to include items:

- on which both Weir (1983) and the majority of the 17 judges from the above study agreed were testing any particular skill (items 2, 4, 9, 10)
- items representative of both "lower" and "higher" order skills, according to Weir (1983) (items 1, 2, 4, 5, 6 and 10 = "lower"; items 3, 7, 8 and 9 = "higher")
- items which the majority of the judges agreed were testing either "higher" or "lower" order skills (items 1, 2, 4, 8 = "lower"; item 9 = "higher")
- items where opinion was more or less evenly divided among the judges as to whether they were testing "higher" or "lower" order skills (items 3, 5, 6, 7 and 10).
- items on which there was more or less agreement among Weir (1983) and judges as to skill **and** level being tested (items 2, 4 and 9).
- pairs of items supposedly testing the same skill (items 5 and 6; 7 and 8).

J. Charles ALDERSON is a senior lecturer in the Department of Linguistics, University of Lancaster. His main academic interests are language testing and reading research, and he has published widely in both areas.

Part One of this paper appeared in Reading in a Foreign Language 6(2).

items supposedly testing Munby skills where both Weir (1983) and the judges agreed on the "level" of the skill (item 2, skill 4, which 78% of judges consider to be "lower"; items 5 and 6, skill 8, which 100% of judges consider to be "lower"; items 3, 7 and 8, skill 10, which 89% of judges considered to be "higher"; item 9, skill 11, which 67% of judges considered to be "higher".)

Table 8 below sets out for each item what Weir (1983) and the judges consider it to be testing.

Table 8: Skills being tested by each item on pilot test

Item No.	(TEEP T012)	Munby skill (Weir, 1983)	Judged Munby skill (n in brackets) (Note 1)	Judged H or L (Note 2)
1	1	7/8	3(7); 2(4); 5(3); etc	13L; 3H
2	2	4	4(13); 3(2); etc	15L; 1L/H
3	5	10	8(7); 10(4); etc	5L; 9H
4	6	2	2(8); 2(4); 12b(2); etc	12L; 3H
5	8	8	8(8); 10(2); etc	9L; 7H
6	9	8	8(7); 9(3); 10(2); etc	8L; 8H
7	10	10	8(6); 10(3); 7(2); etc	8L; 8H
8	11	10	5(5); 12a(3); 10(2); etc	10L; 5H
9	14	11	11(10); 9(3); 8(2); etc	2L; 14H
10	15	4/5	5(8); 4(2); 6(2); 4/5(1); etc	5L; 8H

NOTES

Note 1: etc. = single judges identifying other skills. For full details, see Table 3.

Note 2: Some judges indicated L/H or ?. For full details, see Table 4.

Two students studying English for Academic Purposes at the Institute for English Language Education at the University of Lancaster were invited to participate in the study, and both agreed with interest. P. is Greek, a graduate of Politics, hoping to study International Relations after completing his language course. J. is Colombian, an Economics graduate hoping to study for a Masters degree in Marketing after his language course. Both students were roughly classed by their teachers as "Intermediate". Both had experience of reflecting on how and why they are reading, as part of their EAP course. Both were thought to have sufficient proficiency to be able to self-observe in English. (Experience showed P. to be superior to J. in his ability to do this in English.)

In addition to looking in detail at what the students were able to report about their thoughts when answering the reading test questions, it was decided to experiment with two different forms of self-observation: introspective (concurrent) and retrospective. Accordingly, after briefly explaining the purpose of the study (viz. to investigate how students actually take reading tests, and what they think as they are doing this), J. was asked to take the test on his own, for one hour. He was told that at the end of the hour, he would be interviewed on how he had answered the questions. P., on the other hand, was asked to take the test immediately and to attempt to “think aloud” as he was doing so. He was then subsequently allowed an hour to go through the test on his own, whilst J. was being interviewed. In fact, P. produced a protocol that was more self-observation than “think-aloud” or “stream of consciousness”.

In each case, the session lasted approximately one hour, and was tape-recorded for later transcription (Appendices 1 and 2). Each session began with a practice retrospective report, when the student was asked to look at a reading test taken during the previous week, and to reflect on how and why he had responded as he had. This task presented no difficulty in either case, although it is suspected that both students had already discussed the test in class with a teacher.

After the two individual sessions, the students were brought together for a final discussion of the task. At this time, the interviewer gave the model answers as supplied by Weir (1983), and these were discussed. This session was also tape-recorded for subsequent transcription (Appendix 3).

RESULTS

P.'s introspections were interesting, but interspersed with long silences. He was able to report what he had done during the silences, to some extent, although attempts to get him to be specific were not usually successful. Nevertheless, he provided useful insights into his thought processes. In the final discussion, and during feedback on the answers, P. also supplied interesting comments on how he had gone about answering the questions during the hour he was on his own. Although he made many linguistic mistakes, he appeared to have surprisingly little difficulty in expressing his thoughts. It is still possible, however, that language problems may have made it difficult for him to be more specific on reading difficulties and problems. J., on the other hand, had considerable difficulty expressing his thoughts retrospectively in English. He was encouraged to use his mother tongue, but refused, saying that the value of the experience was to practice and improve his English. Whether more useful results would have been gathered in Spanish is unknowable, but the fact that he was retrospectively may itself have caused the shortness of his responses, and their relative lack of information.

The difference between the two subjects seems to have less to do with whether they were introspecting or retrospecting, and more to do with language ability, and possibly how good they were at reporting their own thoughts. In qualitative research of this kind, it is more important to identify good informants than to find representative informants, and P. appears to have been more informative. Interestingly, in the final discussion P. reports finding the experience enjoyable and "a little amazing" – he subsequently told his teacher he had learned a great deal about reading and taking tests from having introspected. J. reported considerable difficulty, at least partly because of language problems. He claimed that fears of making mistakes in front of a native speaker led him to concentrate more on form than content.

TEST-TAKING VS READING

Both in the warm-up session, and during the test proper, P. claimed that he could understand the texts, but had difficulty with the questions. "I think the questions are much more difficult than the text" (lines 1.20/21) In item 3 in particular he both reports and displays considerable difficulty understanding the question, (lines 1.255/59), a feeling he confirms in the final discussion where he claims he "has difficulty understanding test items, not the passages" (lines 3.51/2). It appears that reading and taking a test are not the same thing. Not only because one can be selective in one's reading if answering a question (although one can clearly also be selective when reading on one's own, the selectivity is self-imposed, not external) but particularly because the test items present both a task one might not have undertaken outside a testing context, and an extra layer if not level of processing difficulty and complexity.

J. claims he reads very differently when taking a test and reading on his own, at home or in the library. In contrast to P., however, he implies that he reads better when taking a test, as he is better able to concentrate and remember. Whereas P. reports considerable negative affective reaction to taking tests, and indeed his protocol shows the effect of a consciousness of time pressure, J. does not appear to have felt pressure or anxiety. This may, of course, be due to the fact that J. is retrospecting.

TEST-TAKING

J.'s approach to the test task appears more suitable than P.'s. P spends some considerable time simply ploughing through the text without reference to the questions, and only a realisation of time pressure leads him to abandon this strategy and refer to the questions. He claims this is his normal behaviour. As a result, however, he fails to gain an overview of both the text as a whole (which causes problems in items 7 and 8) and of the variety and range of difficulty of the items

themselves. He is obliged to become selective in his answering of test questions, by time pressure, but this is done relatively inefficiently. In the final discussion, he reports having learned from the experience that a more useful strategy would be to read through the question first (lines 3.08/9, and 3.19/23). J., on the other hand, begins by skimming rapidly through the passage to gain a general idea of what it contains, and then quickly through the questions, before beginning to answer them systematically. Unlike P., who failed to answer four questions, and whose protocol only refers to 7 questions, J. answers all ten questions and comments on them all. J. appears to be a more efficient test-taker in general (although he still has problems approaching particular items: on item 3 he failed to notice that the instruction directed him to Section I and not the Introduction).

METHOD EFFECT

The detailed discussion of items will show that the way the subjects process both text and items is often conditioned by the test method being employed. This is perhaps clearest when contrasting the reports of the warm-up phase, on a multiple-choice test, with the test proper, which largely employs a short-answer item type. In the case of the multiple choice test, P. in particular seems to proceed by eliminating implausible distractors, before identifying the "correct" response. This strategy results in both correct and incorrect responses. He does, however, have difficulty choosing between two options (in 1.62/69) when they appear to be similar in meaning. In the multiple choice items, he also identifies key words, often unknown to him, amongst the options, and then tries to find them in the text (items 4, 5, 6 lines 1.69/76). Presumably, if P. had not been presented with these words he would not have attempted to search for them in the text. Thus the provision of distractors may cause P. to do things he would not otherwise have done. In all test types, however, a common strategy is to seek to match key words in the question with the same or similar words in the text. This is presumably a strategy confined to test-taking and scanning tasks, and is an example of how reading and answering test questions may be different from each other.

Interestingly, P. points out in the warm-up phase how he has clearly changed his mind and his responses to items 4, 5 and 7 during the test (and in each case, the second attempt resulted in a correct answer). It may be the case that the provision of options in the multiple choice format has actually prompted this change of mind, which might therefore not have occurred in a short answer test. Certainly the existence of options to choose from presents test takers with ideas or alternatives they may not have thought of otherwise. This may have a negative, or confusing, effect, or it may, as possibly in this case, have a facilitating effect.

P. also reports that he does not guess in choosing a response, but rather uses some sort of "logic" (lines 1.109/110). Again, it may be that he would not have used this

“logic” if he had not been taking a multiple choice item. Thus, the “same” question in two different formats may very well involve test-takers using different processes or skills.

On approaching item 8, P. immediately remarks that this will be an easy question, and he points to the dotted lines which indicate both the place for the answer and its likely length (lines 1.358/366). Again in the final discussion P. comments that this item was easy, and typically students enjoy such items. This matching type of question appears to evoke positive affective reactions, at least from P., although in fact, he does not initially find it easy because he has not read the text through, nor gained an overview of its contents (lines 1.363/4). He is able to begin completing the item, however, by employing the above-mentioned test-taking strategy: he tries to identify words from the section headings in the subheadings and the text proper (lines 1.368/377) in order to complete the matching task. J. clearly feels that the test method in item 9 – giving a description of the graph in his own words – involves more than his ability to read and understand the text. He feels he has no difficulty in interpreting the graph, but is very unsure of his ability to express his understanding in words. His mother tongue summary demonstrates this understanding, but his English version, although probably acceptable to a test marker, is somewhat more problematic. Certainly J. is unsure as to its acceptability. More than “pure” reading ability seems to be being tested here. Finally, in item 10, the test method is such that J. does not refer to the text itself, and agrees that it is perfectly possible to answer the question without reading the text. The way the item is constructed – putting sentences into the correct order – leads J. to make judgements about the “logical” organisation of ideas in a sequence or paragraph. Other test methods on the same paragraph would conceivably have involved different thought processes.

ITEM RESULTS

In what follows, I discuss the introspective and retrospective protocols (see Appendices) and compare them with what the test constructor and the item judges believed the items were testing.

Item 1:

This item is intended to test “the understanding of conceptual meaning, e.g. comparison, means, cause, result, purpose” (Weir, 1983, p335, skill 7 and Table 2 above) and the ability to “understand explicitly stated ideas and information” (*loc. cit.*, skill 8). About half of the judges think it tests “understanding relations within the sentence”, although there is considerable difference of opinion. Both the judges and the test constructor are agreed that this is a lower order skill.

In tackling this task, P. reads two full paragraphs in order to “catch the exact meaning”. He then reads three times the sentence containing the target phrase. There is no evidence that he does not understand this phrase – indeed lines 1.224/5 show his identification and understanding of the notion of cause. He verbally supplies a grammatically inaccurate phrase “which the cause is that the multiplicity of factors involved”, and his protocol shows his perception of the need for a relative clause – he is clearly undertaking grammatical analysis of the sentence and possible alternatives, and making some form of judgement about acceptability. He says he would write out the draft answer and then correct it (lines 1.216/8). He admits to not knowing another phrase to replace “because of”, which is perhaps why he resorts to a relative clause. In the final discussion, he does not recognise any of the alternative possibilities suggested in the mark scheme, claiming that he knows the words, but not the phrases in which they appear. He does, however, recognise “due to”, but excuses his failure to produce it on the grounds that it is a “rare” phrase. Despite his discussion of possible responses, he failed to supply a written answer to the item and therefore the test interpreter would presumably conclude that he does not possess the skill supposedly being measured. What he seems to lack is simply an actively available synonym for “because of”.

J. on the other hand, knows the phrase “due to”, recognises it as a synonym of “because of”, and feels that it is a “logical” replacement. He makes no reference to any of the text other than the sentence containing the target word.

Item 2:

This item is intended to test the ability to “understand relations between parts of text through cohesion, e.g. reference” (skill 4, Table 2 above). The majority of judges agree. All are agreed that this is a lower order skill.

P. eventually answers the question correctly, after reading several paragraphs, then re-reading several times a sentence two sentences removed from the “it” of the question. He appears to be trying to understand “the whole meaning” of the paragraph, as well as the specific referent of “it”. He notices a possible parallelism with a previous “it”, which he takes to refer to the same noun phrase (lines 1.247/9). Interestingly, he makes a physical connection between the anaphoric pronoun and other parts of the text. He does not provide evidence of understanding the (correct) answer, and is unable to say why he prefers that answer to an alternative noun phrase: “the other substances such as petrol” (line 1.251). He shows uncertainty about the meaning, and about his answer; he also shows some anxiety because of the time it has taken him to produce a draft answer. Because of time pressure, he moves on to the next item.

J. does not know the meaning of "sniffing", but feels that from the context, especially the text following "sniffing", he has a general idea of its meaning. He analyses the propositional content of the paragraph, noting that the author moves from writing about "drug dependence" in general to specific forms of drugs. He thus infers that "sniffing" is one form of "dependence drug" and then successfully connects "it" to "sniffing".

Item 3:

This item is intended to test the ability to "separate" the essential from the non-essential in text: distinguishing the main idea from supporting detail, e.g., by differentiating the whole from its parts, fact from opinion, statement from example, a proposition from its argument" (loc. cit., skill 10 and Table 2). Only four judges agreed with this, and the majority considered that the item was testing the ability to "understand explicitly stated ideas and information" (skill 8, Table 2). Thus, Weir (1983) considers this a higher order skill, and the majority of judges agree, despite having classed skill 8 as "lower order" (Table 2).

P. has problems understanding the question itself, which he says is a problem in his first language as well. In fact, he misunderstands the task, thinking he is required to write down six key words that indicate the main idea of the paragraph, whereas the test constructor presumably required the first six words of the sentence that best summarises the paragraph.

Although he fails to produce the required answer, and his response would probably be considered incorrect, his protocol indicates that he is trying to identify the main ideas in the paragraph, with some success. Failure on this item does not necessarily indicate inability to grasp the main idea of the paragraph, although it is difficult to evaluate his answer as adequate.

J. identifies the general idea of the second paragraph of the Introduction as being "addiction", and feels that the other sentences are more specific. He thus responds with the first six words of the first sentence of that paragraph. He clearly demonstrates the ability to identify the main idea in text. Unfortunately he has misread the instructions of the item and has referred to the Introductory Section, not to Section II. He would therefore presumably be penalised, and one would infer from his item score that he was unable to identify the main idea, yet his protocol shows clearly that he does have this skill.

Item 4:

This item is intended to test the ability to "deduce the meaning and use of unfamiliar lexical items through understanding word formation and contextual clues" (skill 2, Table 2). Half the judges agree, although there is considerable variety of opinion. Most agree with Weir (1983) that this is a "lower order skill".

P. has difficulties with this item because of lexical problems. He relates “unconvincing” to a word he claims he knows: “convince”, and recognises that “unconvincing” must be in some sense opposite in meaning, but fails to identify the meaning. He does not know what “decisive” means, and so abandons the question. He presumably fails because of lack of lexical knowledge. Although he gives some evidence of the ability to relate words morphologically, this does not help him. Nor is it clear that an ability to guess meaning from context could possibly help if he does not know the meaning of the words whose synonym he is supposed to identify.

Although J. says he does not know exactly the meaning of “inconclusive”, he relates “inconclusion” to “conclusion” and the meaning of “fact”. He decides that “not decisive” and “unconvincing” indicate “we do not have any proof about this”. In fact, he reveals that he knew both words “decisive” and “unconvincing” as they are both similar in his mother tongue, Spanish.

Item 5:

This item is intended to test the ability to “understand explicitly stated ideas and information” (skill 8, Table 2). Half the judges agree, whilst the other half identify a variety of alternative skills. A small majority of judges agrees with Weir (1983) that this item is testing “lower order skills”.

P. fails to respond to this item during the interview, as it appears to be difficult. He prefers to move onto item 6. Afterwards he answers the item correctly, but in the final discussion he fails to comment on the item.

J. answers the item correctly, and says that he only found one sentence in the paragraph that refers to what smokers believe. He appears to have identified the main argument of each sentence, and to have evaluated its relevance to the question, (lines 2.168/70). He is clearly separating the relevant from the irrelevant during his answering process.

Item 6:

This item is again intended to test the ability to “understand explicitly stated ideas and information” (skill 8). Just under half the judges agree with this, the majority identifying a range of alternative skills. Opinion amongst judges is evenly divided as to whether the item is testing “higher or lower order” skills.

On this item, P. has further lexical difficulties, this time with “disease”. However, after reading the question three times, he decides to try to guess the meaning from the context, and subsequently claims he has been successful in doing so (line 1.310). He reads through the section to find the relevant lines, then reads the whole of the paragraph containing the target sentences. He decides to ignore the tables and

figures, possibly because of his unease with such diagrams (see final discussion), possibly because he rightly considers them irrelevant to answering this item. He then re-reads the question and underlines what he sees as the key words in the item. He recognises the link between "connection" in the question and "relationship" in the text and his protocol reveals that he has identified two opposing points of view, for and against a relationship between smoking and disease. He then correctly selects the argument required by the question, and quotes the appropriate piece of text. Finally, he re-reads the text to confirm his response. P.'s answering of this item reveals an extended process of matching, identifying propositions, making judgements about relevance and irrelevance to the task in hand, and judgements about the adequacy of his response.

J. recognises that the proposition of the answer to this item is in contradiction with the previous item (lines 2.178/85), and is clearly demonstrating an ability to compare and contrast ideas in text. He has obviously understood the main idea of each argument, although his protocol does not reveal how he went about reaching that understanding.

Item 7:

This item is intended to test the ability to "separate the essential from the non-essential in text" (skill 10). Only three judges agree with this, and just under half agree that it is testing the ability to "understand explicitly stated ideas and information" (skill 8, Table 2). The judges are evenly split over whether this item tests "higher" or "lower" order skills. Weir (1983) believes it to be testing "higher order" skills.

P. recognises the need to find a mention of "Coronary Heart Disease", but fails to realise that this is contained in a section heading. Instead he searches through the text and fails to find it. He reports not knowing the meaning of "heading", yet correctly guesses that it means "title". However, he then fails to look for such a title in the text. He experiences feelings of panic, probably brought about by his initial failure to find the heading, and his uncertainty with respect to the meaning of the question. He searches somewhat frantically for some piece of text that might help, clutching as it were at straws as he goes through the diagrams and figures in vain. He finally decides to go onto the next item because it seems easier. The effect of negative affect in a test situation is clear here. In the final discussion he reports having found the relevant section once he was alone and no longer under pressure (lines 3.04/5). He blames himself for this, rather than the test-taking experience (lines 3.57/9). His eventual answer was correct.

J. reports having searched for the "title" "Coronary Heart Disease" and then having tried to identify a specific sentence within the relevant sub-section. He reports having rejected the first paragraph of the sub-section as irrelevant because it referred to

statistics (lines 2.207/11) and did not help him answer the question. Interestingly, although it appears possible to answer this item by matching the words “the strongest factor” in the question with the phrase “Probably the most important factor in smoking”, neither subject reports having done this exactly, although J. identifies the first sentence of the third paragraph as providing the required answer “because they are talking about what is the strongest fact that influences in death. ” (line 2.191/2).

Item 8:

This item is also intended to test the ability to “separate the essential from the non-essential in text” (skill 10). Only one judge agrees with this, and there is considerable difference of opinion among judges as to the skill being tested. The highest agreement is among the five judges who believe the item is testing the ability to “understand relations between parts of text by recognising indicators in discourse, especially for introducing, development, transition and conclusion of ideas” (skill 5, Table 2). Most judges disagree with Weir (1983) and feel that the item is testing “lower order” skills.

P.’s processing of the beginning of the item has been referred to above: because he has not skimmed the text or the test, he has failed to gain the overview of both that might have facilitated his answering the question. Nevertheless, he begins successfully to match words in the options with key words in sections of the text, by going rapidly through the sections to gain a general impression of their contents. Once on his own, he manages to complete the item with reportedly little difficulty.

J. answers item 8 correctly by what he reports as a process of “logic”. The text proceeds from the general to the specific. “First the text told me about the drug dependence and after this he talks about more specific topics. In this case, after “Smoking” and after more specific about death and after this, is more specific about different kinds of diseases” (lines 2.237/41). He does not, however, feel that he could have answered the question correctly merely by a process of identifying the logical ordering of sections, without referring to the text. He relates the titles of text sections to their content: “a lot of disease associated with smoking” and the associated tables and graphs. He also seems to have gone through a process of elimination that shows some element of test-wiseness: because he has associated a particular subheading with one Section, that subheading cannot also refer to another Section. “In this case, I can’t associate Section II with c), because I have already answered the question” (lines 2.230/2).

Item 9:

This item is intended to test the ability to “transfer information from one medium to another” (skill 11, Table 2). The majority of judges agree, and fourteen out of sixteen agree with Weir (1983) that this is a “higher order” skill.

P. fails to complete the item during the introspection, but makes interesting comments on it in the final discussion. He did not attempt to answer the question because, as he claims, it is biased against his social science background. He says that he hates diagrams and the like, and that he has difficulty understanding them. Whether this is a background knowledge effect, or a lack of the skill being tested, is impossible to determine. The problem appears to have been compounded, however, by a failure to understand the word "overleaf" in "Figure 21.3 overleaf". As a result, he appears to have wasted time trying to understand a table irrelevant to the question. However, even when given an opportunity to interpret the correct figure, he appears to find this difficult. Whether this represents genuine disadvantage because of his background it is difficult to say, but his affective response to the task is clearly negative and possibly this is an important factor in his failure to respond to the item. Once more, it is hard to conclude from an absence of response to the item that P. lacks the skill supposedly being tested.

J. found item 9 difficult, not because he failed to understand the Figure in question, nor because he was unable to say what the graph indicated about the effects of giving up smoking. He provided a perfectly adequate explanation of this in Spanish. However, he reported great difficulty in expressing these ideas in English, and considered this item the most difficult of all, for this reason. He is unsure whether his response is adequate, and shows considerable interest in the interviewer's reaction to it, during the final discussion (lines 2.262/2). He does not comment on P.'s allegation that because of his (J.'s) Economics background, he is at an advantage in interpreting the graph, but he certainly appears to find the interpretation straightforward.

Item 10:

The item is intended to test both the ability to "understand relations between parts of text through cohesion devices" and the ability to "understand relations between parts of text by recognising indicators in discourse" (skills 4 and 5, Table 2). Most judges agree with this. More believe that the item is testing "higher order" skills than "lower order" skills, and in this they tend to disagree with Weir (1983).

P. did not attempt this item.

J. orders the sentences in this item in terms of their "logical" relationship to each other. "OK, I have to organise my ideas". He identifies those words in each sentence which indicate the relationship of one sentence to another — e.g. the referent of "his", the reporting of the experimental results for subjects in numerical order: "Subjects one, two, three", "the next three", "number 8"; and so on (lines 2.305/11). He recognises, however, that it is perfectly possible to answer item 10 without reference to the text, and that therefore it tests his ability to understand the question, and not the text.

CONCLUSION

The main point to emerge from the discussion of test method effect is that there are clearly certain skills or processes involved in and required by answering test items in a particular format that are not necessarily specified by the test writer. Weir (1983) does not refer to the matching process, to the elimination of alternatives on some rational basis, which presumably involves making judgements about sense and plausibility, he does not refer to the ability to express one's ideas in writing, nor to several other processes discussed above. Munby's (1978) taxonomy does not refer to these skills either. Yet they are clearly important components in answering these reading test questions.

It is furthermore reasonably clear that the skills involved in answering tests in general are somewhat different from and either additional or alternative to the skills that the test constructors intend to focus on. This is despite the fact that both the subjects reported finding the tasks interesting and enjoyable, and despite J.'s (naive) view that the tests are good tests of reading. Face validity they may have, but to what extent do they have content validity in the sense that they test what their constructors claim they are testing?

What appears to emerge from the detailed comparison of performances on individual items with what various parties believe the items to be testing is: firstly, that the two subjects sometimes approach the items in different ways, and different processes are involved. This is particularly the case when one subject does not know particular lexical items. Secondly, what appears to be being tested by an item does not always match the beliefs of judges and test constructors (see especially items 3, 4, 5, 6, 8, 9). Thirdly, in several cases, the subjects appear to be using various skills to respond to an item, rather one single or even one main skill. Fourthly, a subject is occasionally seen to demonstrate the skill supposedly required by an item, and yet fails to answer the item correctly. Thus, fifthly, it would be erroneous to conclude from an incorrect response to an item that the subject did not have the skill in question (especially true for item 3). Sixthly, conversely, that subjects do on occasion appear to respond correctly to items without displaying the skill(s) in question. This is especially true in those items which require test-taking skills like matching.

Obviously the pilot study reported is very limited in scope, and there are limitations in the way the interviews were carried out. In particular, it was often difficult for the interviewer to probe carefully into the process of answering individual items, and only in retrospect was he aware that certain areas should have been explored more carefully and systematically at the time. This is perhaps the inevitable result of a procedure which is reactive to an ongoing process, rather than the application of a pre-determined instrument. It was only on analysing the transcripts that certain

questions or issues became apparent. Ideally, the analyst would then go back to the informants, as soon as possible after the event, with supplementary questions and requests for confirmation of interpretations.

Nevertheless, despite the limitations of the study, it is felt that there are findings that are worth stressing. The evidence for the existence of separable skills is not overwhelming, and the test-taking process (and therefore, by inference, at least part of the reading process) probably involves the simultaneous and variable use of different, and overlapping "skills". The division of skills into "higher" and "lower" orders, however tempting, does not seem to be justified in practice, at least to the extent that it is possible to say that subjects are using "higher" order skills which imply necessarily the ability to use "lower" order skills. The relationship between skill "level" or "complexity" and item "difficulty" is far from clear, but is certainly not simple. In the case of these subjects, difficulty was often associated with knowledge of particular lexical items, with particular test methods, or with the involvement of macro skills like writing as well as reading.

Needless to say, further research is necessary and careful thought needs to go into designing elicitation procedures in order to illuminate the research questions more precisely than was possible in this case. There are, however, important implications of this research, not only for theories of reading and their operationalisation in tests of reading, but also for the process of test construction and validation. What a test of reading tests is not simply what its constructors say it tests, nor what a set of judges considers it to test. It must surely and crucially relate to what happens inside a test-taker's head when he or she responds to an item. Finding out that information, and discovering how generalisable the results are, is a neglected but important research endeavour.

APPENDIX 1

Note: Line numbers given here are those from the original manuscript, and are not accurate in terms of the printed text. This is to preserve the accuracy of references in the body of the text to sections of the appendices.

INTROSPECTIONS BY P.

JCA: OK Because you've not done any of this before, I thought it might be useful to go through this first. Remember taking this test?

P: Yes, yes

JCA: It's a reading comprehension test. I thought I'd just ask you one or two questions about it. This first one

P: Yes

JCA: It says the extract –the text is about wrestling. Do you remember that?

1.10 **P:** Yes

JCA: When did you take this test? Was it last week?

P: Yes last week

JCA: How did you find it?

P: The text? I found it rather difficult, but not so difficult. The questions, it's another area. The test, individually, it was difficult for me, not to understand, to catch the meaning, but to catch special points in this text.

JCA: OK, so although you understood the text, you had difficulty with some of the questions?

1.20 **P:** No, I think that the questions are much more difficult than the text.

JCA: OK, well, let's have a look at this first one. It says the extract is taken from these four areas, and you put a tick against two: local history book or guide book.

P: No, after, I put this one: Guide book

JCA: I see, and then you changed your mind?

P: No, I didn't change my mind, I corrected it when we revised it (in class).

JCA: Why did you put "Guidebook"?

1.30 **P:** I put "Guidebook" as I remember because I think that the text wasn't too much to be from a leaflet book, it's too big for a leaflet, I think. I didn't put "history" because I thought that it's too, that it doesn't give so many information to be from a history. I didn't put "encyclopaedia" because, as I know they give very small information. That's why I preferred to put Guide.

JCA: OK Did you go through each of these options and decide they were wrong before choosing the right one?

P: Yes, I did not choose the right before, I put this, yes.

1.40 **JCA:** OK, that's interesting (*JCA misinterprets. P means "I got it wrong"*)

JCA: Second one, then: where does it come in the longer passage, the end, the middle or the start?

P: I said in the middle

JCA: Why did you say that?

P: Because here it starts with “But wrestling” and it makes me think that if there is “but”, there is something else before and you never start a sentence, a paragraph, not a paragraph. You don’t start a theme with “but”. You refer to another thing: maybe sports in general. And it doesn’t look to finish here, the whole text.

1.50

JCA: What makes you think that?

P: I don’t know exactly. Something like a feeling. I can’t say that there isn’t a conclusion. But there isn’t a conclusion. Because I think that there isn’t here. It looks like to be a part, from something to say about sports or about something in this area, in Lake District. I think that it would be conclusion to say other things after that. There isn’t a conclusion, there is a conclusion only for this part, for this analysis of this sport, of wrestling.

1.60

JCA: OK, good. And what about the third one?

P: Ah, these are the difficult questions. It was a little puzzle. I have to say that I was tired to do this “threw his opponent to the ground”. This must be . . . I don’t remember exactly the place, but in this question; I tried to find from the words that are in this text because I didn’t catch the meaning exactly, which words, which meaning is most close in this phrase, because I didn’t catch the meaning exactly. And from the text and from here, I tried to put the answer that is more closer from the words here and from the words to the text. I did the same thing and in the fourth question and the fifth question. In this one, in the sixth, I couldn’t understand the meaning.

1.70

JCA: Of the question or of the text?

P: Of the question and of the text

JCA: OK

P: In this I did the same thing as in the fourth, third and the fifth

JCA: So you were actually looking for these words, were you?

P: Yes, I was looking for these words because I didn’t know what “embroidered shorts” is, for instance. In this one (*reads stem of Number 8 to self*) it wasn’t well, it was a little smarter than the others, I would say.

1.80

JCA: Ah ha, Number 8?

P: I remember what I did in this exactly. Ah, here it is. (*Reads from passage*) “Wrestling in the dales is a sport for the man who knows . . .” I think that here we can see two answers in general, but I didn’t catch exactly what you preferred this one. There is rather a small difference between this and this one

JCA: a) and b)? Ah ha.

1.90 **P:** But I grasped the sense (*Reads from passage*) “Wrestling in the dales is a sport for the man who knows the game . . .” I put this because I thought that this must be the different from this. I tried to catch the meaning, as I remember, I don’t remember very well, I tried to catch the meaning from the words. I know the meaning, I know exactly what this means. I knew the words and this one, but I think that the difference between a) and b) is very very light.

JCA: So did you make a guess in the end?

1.100 **P:** No, I didn’t guess. (*Reads silently for 18 seconds*) I found that this argument was stronger than the second one, that the person who wrote the questionnaire maybe wanted this one. I thought that this is a little stronger than this one. That’s why I did it. I didn’t guess in this one.

JCA: OK, and in the last one?

P: In the last one. This was my opinion. I don’t know if I, I don’t remember if I catch it from here but it was my opinion, that there isn’t any mention (*reads from question 9*) “to persuade people” “to try and prevent wrestling and fell running”. The same thing for the. .

JCA: So it was logical?

1.110 **P:** Yes, it was logical. I don’t know if you want some extra information.

JCA: No, that’s fine. That was very interesting, actually. Did you enjoy that test, or not?

1.120 **P:** This one, yes. Finally, when I saw the mark – not the mark, because I am against the mark – when I saw that finally I guessed right, it was something like a puzzle. Something like this. This one I really enjoyed it. It was rather difficult, I think. These questions and there were a big distinguish between these questions and the text. And there were questions that were too difficult, as this, and very easy, as this, and this one, and this. Eight was easy for me. Nine was easy. But 6, 5, 4 and 3 were difficult. I left the fourth one – you can see here (*points to erased pencil mark*), I changed my mind, and in the fifth one and the seventh. I changed my mind. Can

you see that? And I took the wrong answers and I changed my mind. It wasn't guess. I can't say that. It was and guess but it was and logical and it was something between them. I could be more specific if I said that I tried to find the words in the text with all this that I understand.

1.130 **JCA:** Fine, thanks very much. Let me now move on to this. This is the test that I want you to take. That is the passage, and these are the ten questions. What I'm going to do is ask you to take it as a test, any way that you would normally do it, and I'm going to try to get you to think aloud. I'll ask you questions. I'll let you begin, though.

P: To read it aloud?

JCA: No, no, you just read it, like a test.

P: To read it first, and after, when we will do this?

1.140 **JCA:** Is that what you would normally do when you take a test? When you take a test, normally, what do you do? Do you read the passage first or do you read the questions first?

P: No, I don't read the questions. Never. I don't know why. I read first the passage.

JCA: OK, fine.

P: Do you want to say what I exactly do?

JCA: No, you just read the passage and then . . .

(Reads for 6 minutes 30 seconds)

P: Do you want to know exactly what I do now, because it's too much. We have only an hour and now I will look at the questions.

1.150 **JCA:** OK, fine, sure. Before you begin to look at the questions, could I just ask you about the text. How do you find it?

P: Not so difficult, not so easy. Exactly in the middle. More easy than the other one (the wrestling one).

JCA: Did you find the text interesting?

P: Yes.

JCA: Why?

P: Because I smoke.

JCA: Ah ha. Do you know anything about this topic, have you read anything about it before?

1.160 P: No.

JCA: You say you found it interesting because you smoke. What did you find interesting?

P: I found interesting because it had something that I did not know in this text. I don't use to know text about smoking because I don't think that there is something interesting in this, but this I have to read, and I found a few things that they are interesting. Much more than the other one about wrestling.

JCA: Is this the sort of text that you would not normally read, in other words? This topic.

1.170 P: No. I avoid to read this topic, not because I am afraid, because I take such a long time to read in general and is from the I think I pass from the magazine or newspaper, I don't read them.

JCA: Where do you think this comes from? From a magazine or a newspaper?

P: No. I don't think. It is more a scientific but it isn't so scientific, but as I see there are sections, it's a research, maybe.

JCA: Why do you think that?

1.180 P: Because there are diagrams, statistics, here, here.

JCA: You said "not so scientific". You said "scientific but not so scientific".

P: Yes I said "not so scientific" because as I think if it was more scientific it would be referred to the medical and it has to have special vocabulary, but I didn't mark any special vocabulary here. Maybe it happened from, I don't know, from a student's notes, it's impossible, from a factory of smoking, from one local doctor or authorities. I don't know, but it isn't so scientific, I didn't think, but I don't know. It looks like could be scientific but I don't know.

1.190

JCA: OK Fine, well if you want to start looking at the test questions now, and try and answer them.

P: I tried to change, to go to the questions because I thought that if I am writing a test, I should not have a long time, it is better to know the questions. That's why. And I don't need all the text. . . . *(Reads for 8 seconds)*

JCA: What are you looking for?

1.200 P: I am looking for the first question “Give a phrase that could replace “because of” in line 3 of the Introduction to the reading passage”.

JCA: So you look for the “because of”?

P: Yes, I am looking for the phrase “because of”. Now I am going to read again, until . . . the first two paragraphs (*reads silently for 30 seconds*)

I don’t know the word “addiction” but I guess this word from this text, I am not asking you, but I don’t know exactly, but I guess. (*reads silently for 36 seconds*). You want what I am thinking or the answer?

JCA: What you’re thinking first, and then the answer.

1.210 P: Yes, I thought . . . I read this passage and I tried to catch the meaning exactly, more deeply than the first time I read it. And now read three times, only this sentence, which contains “because of”. I thought that it isn’t possible to replace by another one, because I don’t know another one but I think a phrase . . . (*reads question again*) “Give a phrase”. Phrase. I give an answer in my mind. If I will write it, I’ll write the answer to see if it is understandable and correct, this one, to make it correct from the wrong, from the errors.

JCA: What would you write?

1.220 P: (*Reads passage*) “Addiction is a complex matter” . . . I forgot it. (*reads silently and thinks for 41 seconds*) Which the cause is that the “multiplicity of factors involved”.

JCA: OK, why do you think it’s that?

P: Because I thought that in this word “because of” there is the meaning of the cause, I tried to make a sentence with the cause and I thought that it isn’t possible to replace a sentence in another sentence in another sentence, like this, but only if this would be a relative clause sentence. That’s why I did a sentence with “which”. (*Reads silently for 6 seconds*)

1.230 The second one. (*Reads silently for 41 seconds*). Now, in the second I’m going to read all this passage which begins with “Dependence”, to catch the meaning until “it” or . . . (*reads silently for 160 seconds. Points to part of text, and then keeps finger on it somewhere around “thousands” (middle of same line as “it”)*).

1.240 I think that when . . . I read this one one time and after that I re-read four or five times the sentence from “Sniffing” to “loss of coordination” and I tried to read the first time fluently, to understand what does it mean, and very fast, and the other time to catch the meaning, I read three or four

times this, before answer and I'm not sure, but it's a test and I have to hurry. This is my problem always and this "it" must refer to "sniffing or solvent abuse".

JCA: Why?

P: From the meaning, I don't catch exactly the meaning but because the previous "it" it refers to this one and from the meaning I conclude that it maybe refers, too. I am sure that it refers to the same as the previous "it" but I don't know exactly if it refers to "sniffing or solvent abuse" – this, that it's one part – or to "the other substances such as petrol", but I feel – the feeling that I get, not from guessing, that it has refer in "sniffing or solvent abuse".

JCA: OK, Fine

P: (*reads silently for 29 seconds*) I read this, the third question, about two or three times because I didn't understand exactly the meaning, not from the words but that happens in my language, when I read something in Greece, that happens normally to everyone. Now I am going to read this one, the first paragraph, because I want to know what it says. And now I write the most principal words here.

(*Reads for 118 seconds and underlines words in text: tobacco, factors, research, children, parents, conflicting, differences in intelligence, academic success*)

I underlined a few words as I read it, as I was reading this text, and from these words now I am going to choose the words that it refers better in the meaning. I have to write six words.

P: (*Reads silently for 97 seconds. Ticks the words, and appears to be counting them.*)

"Tobacco is the first. I choose this one because is the principal meaning, (*reads from text*). After that, the second is "research", it's the way that we see the tobacco, we see the tobacco from this aspect, from the aspect of the research in this paragraph.

The third one, I read first "parents" and second "children" because the parents is the problem and the children is the bad effect. And this one four, five is intelligence. That it's "intelligence" is the evidence of this research but here I don't know, this is "academic success", can we use it as one word? or as two words.

JCA: Two words, I guess

P: Two words. I'll put "success" only because I'll leave . . . success, you can have success in life, more than academic. This for the third one. Do you want to . . . ?

JCA: No, that's OK.

P: (*Reads silently for 24 seconds*) I read the fourth question. I don't know what this word means exactly ("unconvincing") and "not decisive", I don't know what this means. "Decisive", I never met this word. "Unconvinced", it's "unconvincing". I know "convince" but now I'm trying to remember what "convince" means, it must be the opposite.

1.290

(*Reads silently for 10 seconds*) I'll leave the fourth one because I've no time, I'm going to go to the fifth one.

(*Reads silently for 15 seconds*)

1.300

Now I think that the questions are going to get more difficult. That's why I am going to the sixth one, because I have no time, I want to see if the other one is more difficult, and I'll come back after.

(*Reads silently for 37 seconds, turns pages*)

Here, the sixth question, I don't understand what "disease" means. And I read this question three times, now I hope to find the meaning from the text but I think here that it starts from this point and maybe I'll catch the meaning from the text.

(*Reads silently for 36 seconds, turning pages*)

I've found where is the sixth (line) but I'm going to read from the first.

(*Reads silently for 24 seconds*)

1.310

I catch the meaning of "disease" from the first time I read it, here.

(*Reads silently for 20 seconds*)

I'm not going to look at this diagram, and the second one, the "Table 21.3" and the "Figure 21.3", I go right to the text.

(*Reads silently for 81 seconds*)

1.320

I read this, understand this and I am going back to the question to see exactly what it means, because I can't remember (*Goes back to the question and underlines words in question: "connection", "disease" and "cigarette smoking", taking 10 seconds.*)

I underlined "connection", "disease" and "cigarette smoking".

(*Reads silently for 67 seconds*).

JCA: What are you thinking about?

P: (*Pauses 6 seconds*) Now I was a little confused here because from "Their main argument" against relationship I was a little confused, but as I see the word "relationship" I understand that it isn't here, the argument that

- 1.330 I was looking for, but I turn to the other sentence, and this (*reads text aloud but unintelligibly, until*) “as people”, it is this one: “as people stop smoking so the incidence of disease and morality (sic) diminishes’.
(*Reads silently for 8 seconds*). This one.

JCA: OK.

P: And I go onto to seven.

(*Reads question for 25 seconds*)

I’m reading Question Seven and I’m going to find exactly where is he says “Coronary heart disease”

(*Reads silently for 44 seconds, from the beginning of Section IV, apparently not looking for a heading*)

- 1.340 I didn’t catch exactly the meaning from the question 7. I don’t know the words, but I can’t understand exactly.

(*Re-reads first paragraph. Then back to question. Then Table 21.3 on page 5. This takes 75 seconds*)

Although I know the words here they seem to have another meaning in Question 7. That’s why I can’t catch exactly the meaning. I don’t know “heading” but I suppose it must be the title. I guessed. And I know “factor” but I think it must be different here as well. Probably? A problem??

(*Reads silently for 26 seconds*)

- 1.350 Now really I think. Do you want to say anything? Now I really feel panic, because I left two questions here, I don’t understand this and I am afraid to go to the other, that’s why I think, and I’m not sure that, I lose my time here but I don’t take the decision to go to the other. Although I try, I think all these things, all this time, I try to catch the meaning and I’m looking to the diagram, to the Figures here, and to the words, I was looking to catch the meaning, but as I cannot find anything, I’ve got to go to Question 8, because it seems to me that it may be much more easier. (*Points to dotted lines to be filled in*).

(*Reads silently for 39 seconds, at one point turning the page to check he has reached the end of the question*)

- 1.360 Now there’s a problem again, because here I don’t know these Sections, because I didn’t, remember I didn’t look, read it very carefully but I am going to do this because it’s something like a puzzle, it must be easy for me, that’s why I am going through the Sections very fast.

(*Reads silently for 28 seconds*)

I’ll go to the first Section, the Drug Dependence. This title must belong to the first Section because I can see that it starts with this word . . . no, it doesn’t start but there is in the fifth line. When I was in the fifth line and I saw this word, I realised that this word must be somewhere around here, and I saw that there is in the fifth line “psychological or emotional

- 1.370

dependence" here and there is in the other that says "Physical dependence" and if I can see the word, I see here "dependence", there isn't this word in this and that, and I put the Section I here, Drug Dependence. We have to stop? OK.

APPENDIX 2

TRANSCRIPTION OF INTROSPECTIONS BY J.

JCA: Before we begin that test, can I show this one to you, to give you a bit of practice. OK? Do you remember taking this test, about wrestling in Cumbria? Do you remember doing this? You did it last week, I think? And you discussed your answers with the teacher?

J: Yes.

JCA: How did you do? Were you happy with your performance? Was it difficult? Was it easy?

J: Not easy. We found some words that I don't know the meaning of, but the majority of the text, it is possible to understand, for me, it is possible to answer the different questions, to understand not all the words, but the principal words of the text.

JCA: When you answered this, did you go through it question by question, or did you read the text first?

J: I read the text first, and after I answered the questions.

JCA: OK. The first question: the extract is taken from . . . and it gives four text types, and you put down "local history book". Why?

J: Because the tourist leaflet is impossible because it is too little, it is something very concrete, a different aspect, very concrete, a long paragraph.

JCA: This is too long to be a tourist leaflet?

J: Yes, it is possible for being something, you need this for some special place, I don't like to read this, it is possible that I don't know about this. And "A Guidebook to the Lake District"?

(Reads silently for 12 seconds) No, no because . . . *(pauses for 10 seconds)*
J: Could be but I think that not, because about the Guide of the Lake District is information that give the book could be different, because in this case speak about something about the sport of the Lake District because other people know the sport of the Lake District. It is possible that it has data about the population, about the place, about the weather, about the restaurant.

JCA: OK. Now the second one. “Where does this extract come in the longer passage”. You put there: “in the middle”. Why did you put that?

2.40 **J:** The start not because the first sentence is illogical. If you want to start a paragraph you can’t put this sentence. “At the end”, no, because you want to have some conclusion about the text. In this case you don’t have any conclusion, you leave open the paragraph and you don’t have any conclusion about this paragraph.

JCA: OK. What about the third one?

J: I think this and this is possible. But the first is necessary; “to throw his opponent to the ground”, (for the second “must be on top of him on the ground”). I think only in this possibility. I think both.

2.50 **JCA:** You think both a) and b) are possible?

J: Yes, a) and b), but put b). This (c) is not possible because is only one of the parts of the sport. Only one of the components.

JCA: Ah ha, that’s c). What about number four? How did you answer that?

J: I couldn’t understand well what is the meaning of the words but about with the text, with the idea I tried to guess the answer.

2.60 **JCA:** Ah ha, you don’t know what “embroidered material” means, but you got the question right. How did you do that?

J: Because about the meaning, because about the words, about the sentence. I don’t know exactly the meaning of the words.

JCA: Did you look for the word “embroidered”?

J: Yes.

JCA: And then you mean you found it?

J: Yes

JCA: OK fine. Now compared with that test, how was this one?

J: Quite easy. Easier than this.

JCA: Really?

2.70 **J:** Yes, I could understand most of the words, and the idea, I think, I understood the ideas, the principal ideas.

JCA: In this text about smoking?

J: I like it.

JCA: You like it? Why?

J: In English when you read and you can understand easy, you like the text, because if you go to the library and you will start to read the text and you need to go to the dictionary all the time, you don't like, it is difficult. Very useful sometime, but quite difficult sometime, you feel bad and you don't like this text, and it is possible that you learn more with the other text than this. But with this kind of text, you feel that your English has improved.

JCA: OK, good. Did you find it interesting?

J: Yes.

JCA: Why?

J: Because they are talking about the life that interests the people. Is a social and interesting topic.

JCA: Do you smoke?

J: Yes, but not a lot.

JCA: Did this text contain information that you knew already? . . . Have you read about smoking before?

J: No, because I know about this but is a general knowledge, not a scientific knowledge. I know about the drug, about the addiction, about the smoking, about . . . I know about the disease, but these are general knowledge, not scientific knowledge.

JCA: Why do you refer to scientific knowledge?

J: About the relationship between smoking and disease, this kind of knowledge, I didn't know already, only general knowledge.

JCA: Ah ha. What sort of person is this written for, do you think? Who is the reader of this text?

J: I think it is a journalist?

JCA: A journalist. Why do you say that?

J: Because is no scientific text, is more a research text where you have some information, you have researched a lot of information, and after this you put this information on this paper. You want to let other people know about this information.

JCA: You don't think it's scientific?

J: No.

JCA: Why not?

2.110 **J:** I think is a person that really knows this topic, but is not a scientific reading passage. Why no? Because the vocabulary is common. Words that everyone knows, not scientific words.

JCA: OK. When you started doing the test, what did you do first?

J: Look at all the text, I tried to read quickly all the answers

JCA: So did you read all the text first?

J: Yes.

JCA: You read it all, from page 1 to page 7? You read everything and then answered the questions?

2.120 **J:** Yes, yes, but after then, before then, sorry I read quickly the answers. When I have this, after I, before I read quickly: what about the text? and after this I start to read. First I read it and after that I answer.

JCA: Right, OK. So, Question 1 then? . . . You put "due to". Why?

J: Because when I read the sentence in my head, it is logical, the meaning I change "because of" and "due to". I read the text and I think it is logical, the meaning does not change. I think this is possible.

JCA: OK. Was that question hard?

J: No, no. I don't know if it is well or not.

2.130 **JCA:** We'll talk about that later! OK. What about the next one?

J: The next one? I don't know what is the meaning of, exactly the meaning of this word ("sniffing"), but for the context, for the explication after this word. I don't know what that word tries to show me, to say me, but I don't know what is exactly the meaning of the word but I can explain the idea, in this case.

JCA: OK. How do you know that it refers to "sniffing"?

J: Because is only the one possibility. Here he start to talk about the dependence and after dependence he talks about different kinds of dependence, and after this, the "it" is only the possibility to "sniffing".

2.140 “Dependence” is more general than this. I don’t know if you understand? Is more general than “sniffing”, only this possibility, a specific kind of dependence drug.

JCA: OK Number 3.

J: Is . . . (*points to second paragraph of Introduction:* “Addiction is a complex matter”)

JCA: Uh Huh. Why did you think that?

2.150 **J:** The general idea of the text is about addiction, the drug, the smokers, alcoholic, the different addiction, and I think this sentence has the more general idea about the text, because the other sentences are more specific, and only study one specific topic. For instance, “The initial problem concerns what is meant by . . .” is a more specific topic, one topic, not a general idea of the context.

JCA: OK. Number four

2.160 **J:** Ah yes. I don’t know exactly what is the meaning of this word, but I think “inconclusion” comes from “conclusion”, yes, and a conclusion is something fact, but in this case “not decisive”, “unconvincing” and “inconclusion” is . . . we don’t have any proof about this.

JCA: Did you know the word “decisive”?

J: Yes, it is the same word in Spanish.

JCA: And “unconvincing”?

J: It is similar to Spanish.

JCA: OK, what about 5?

2.170 **J:** 5? In the paragraph, is only this sentence that summarises me the cause of the relationship, of this opinion of the cigarette. Is only this sentence, because the other sentences (*looks for paragraph*) does not have any explanation about what the smokers say or believe about the disease. Then this is the only sentence possible.

JCA: So that sentence is the only possible sentence because that’s the only one where the author has. .

J: Yes, in the other sentences, I don’t have any idea about what the smokers say about the smoking habit, about the relationship between smoking and disease.

JCA: OK. What about 6?

2.180 **J:** Is a refuse (contradiction?) with relation the before sentence, because in this case they are showing that really there are a relationship between disease, mortality disease, and refer to smoking cigarette.

JCA: And the previous sentence?

J: This is a refuse, because here the people say, OK, we don't have any relationship and here they are showing that in fact there are a relationship.

JCA: So 6 is the opposite of 5? OK.

J: 7. This is a long explanation, but I only put the first sentence because . . . I think that the sentence is very clear. The explanation about this is very clear.

2.190 **JCA:** Why is it very clear to you?

J: Because they are talking about what is the strongest fact that influences in death, organ in the body of the human? This is all the explanation about this.

JCA: How did you find the right answer?

J: About . . . ? why? Because in this case . . . from the title.

JCA: OK so you found the title. Did you look for any of the words in the question?

JCA: Yes, this.

J: You looked for "Coronary heart disease". And then?

2.200 **J:** Then I am going to the text and I chose a specific topic about this.

JCA: Which topic?

J: That showed me about "Coronary heart disease". And after this I read it.

JCA: And when you were reading, what were you doing? Were you looking for anything?

2.210 **J:** I tried to concentrate in answering the question. I started reading, for instance in the first paragraph, the idea is about some statistics about the death in Britain and USA. I think, I know, that is not useful for me if I want to answer the question.

JCA: OK and so?

J: And after I found the answer of the question.

JCA: OK. Fine. Number 8, then?

J: Number 8. This is the general idea, the main idea. Number c) is very logical because it is a special part of the text, it talks me about the diseases. The “Smoking” section was very clear, like show me the different kind of smoking. And Number III is very clear too. I think this dividing is very clear, about how the text is divided, the different sections.

2.220

JCA: OK. You said that Section II is “Smoking”. Why is that not “Diseases Associated with Smoking”?

J: I think . . . I put first this, Number Four, Number c, because we have a lot of disease associated with smoking.

JCA: On page 5. Ah ha

J: This Table, this graphic, we have a lot of relationship between disease and smoking. And after this, I put Smoking Section II. I answered first this paragraph.

JCA: So you did c) Section IV, first. Yes?

2.230 **J:** And after this I think Section II – Smoking– No this, in this case I can’t associate Section II with c), because I have already answered the question.

JCA: OK, but why not d) instead? Section III? Why not have “Smoking” the title for Section III?

J: (*reads for 15 seconds*) I think in this case, this is the first more general, and when the text is coming, is more specific. Do you understand me? And if you think . . . First the text told me about the drug dependence and after this he talks about more specific topics. In this case, after “Smoking” and after more specific about death and after this, is more specific about different kinds of diseases, and this is the reason why d) is III

2.240

JCA: OK That’s interesting. Could you have answered this without referring to the text?

J: I don’t understand.

JCA: Could you have put your answers there, I, II, III, IV, without looking at the text?

J: No, no.

JCA: So you do that, and then try and answer?

2.250 **J:** I am not sure. But is possible. Really we have a logical ordering. I am not sure.

JCA: Yes, well you said that there is a logical order which is general, and then specific, and you can see that here, with these words, without looking at the text.

J: Is possible, yes.

JCA: OK. Good. What about number 9, then?

J: (*reads question for 21 seconds*) I understand what is the meaning of this figure, but it was very difficult for me to write all the ideas that I want to express, because I show, I look, I understand but when I want to write, it is more difficult for me. I write something like . . . I don't whether it is good what I put.

JCA: Would you have preferred a choice of questions, like this one, where you just tick your answers a), b) c)?

J: OK I think that this is more easy than this, but in this case you don't have any research about if your English has improving or not. Really, this kind of question you know because . . . In this case, OK, OK I have nine out of nine, but I only, I need to choice, but in this case I need to write, to think more, I need to . . . it is more difficult.

JCA: Are you happy if we test your ability to think?

J: Not all the test! I think no. Much of the test, because for instance my case I need to present an exam for my Masters. All my test is very difficult, I prefer this.

JCA: Multiple choice?

J: Yes. But if I want to study, I want to do homework, for me it is better to write, because I don't have problem with tell Ann, please, this is my homework, please I have any mistake or mistake, could you please tell me if I have? And for me doesn't matter if I have. She is my teacher in English. I do not have any problem, but when you have a test you feel nervous, is possible that something in your life, for instance in my case the Masters is very important for my life, and if I can get or not get, is something very strong for me, for my life.

JCA: But if this question is supposed to test your ability to understand, and you say you understand, but you cannot express your understanding. Is that fair?

J: I think, is sometimes, no, in this case, I need to put some special words, I think, OK I understand this, I think in Spanish, I need to put this in

2.290 **J:** Spanish because it is the more correct sentence. But when I want to put this in English, is impossible for me, in this case, to change the way, and to put in other words?

JCA: OK, you tell me now in Spanish the answer to this question.

J: OK.

JCA: Que dice la figura a sobre el efecto de . . . ?

J: La rata de mortalidad por fumar cigarillos es menor entre mas temprano deja la gente de fumar.

JCA: En que sentido temprano?

2.300 **J:** La edad. Entre mas temprano deje el habito de fumar, verdad? But I don't know how to explain it. Between . . . I cannot. I don't know the words.

JCA: OK OK Fine. What about number 10. How was that?

J: OK When you want to order a specific paragraph, you need to put . . . OK, I have to organise my ideas. For instance, Number 1. It says that, it talks about some "studies", only some "studies", about the relation between "smoking and carbon monoxide levels". In number 2 we have a name of a person who make some, each study. In number 3 what refers "his" is only this person, Doctor. And number 4 . . . number 3 give me . . . he is going to prepare a special research about this kind of disease and number 4 he give me a list of the person, what kind of research they are want to make. Number 5 is logical: "Subjects one, two, three" because we have eight. Number 6, "the next three" and number 7, "number 8".

2.310

JCA: OK. Could you have done this, number 10, without reading the text?

J: Yes, yes. Sure.

JCA: Do you think this tests your ability to understand the text or the question?

2.320 **J:** The question.

JCA: OK. How well do you think you did on that test?

J: How? Sorry.

JCA: How well did you do? Did you get them all right? 50% 2%

J: No. I think it depends why you pretend with the test. If you want to have an idea about writing, about how a person expresses an idea, about how a person understands the different text, it depends, what is the purpose of the test, because it is different for the reading, for the listening, for the writing.

JCA: But how many answers do you think you got correct?

2.330 **J:** (Pause) Eight

JCA: OK. Which are the two you got wrong?

J: Well, I think this (Number 9). I got confused. It is possible not to understand.

JCA: You didn't understand, or I won't understand?

J: No, you. Not me! I can guess. . . and Number 4, I'm not sure . . . I think . . . I am more sure in this question than this.

JCA: OK. You are more sure of number 4 than number 9. But which of the two was the most difficult question?

J: Number 9, for me, not for understanding, for writing.

2.340 **JCA:** For writing. Forget number 9, leave aside number 9 because of the writing. Which of the others was the most difficult?

J: Perhaps number 1 or number 4.

JCA: Why?

J: Because in this case you need to put something about your knowledge of the text will give you the answer.

JCA: It depends upon what words you know? Ah ha. OK. Do you think this is a good test of reading?

J: Yes.

JCA: Why?

2.350 **J:** I think the test has different parts, where it is possible to detect if really the person understands the main idea about the text. I think. I am sure that the person who knows nothing about English is impossible to do this test. Sometime I think that this is easy because I understand, it is easy or difficult, whereas to write and to answer this question is possible this person can be able to answer this test. Sometimes you have this kind of question, because normal when you are learning English, when you read it is

2.360 difficult. If it is easy, you can say you know all about English, I can speak English: I have learned English. But while you are learning English, all the tests should be difficult, because all new words, different structure, new grammar.

JCA: When you were taking this test, were you reading it in the way you normally read English?

J: No, it is different. In a test I have pressure, it's a special research, a special exercise, you put more encouragement because you are measuring my English. It is quite difficult.

JCA: So you are more careful?

J: Yes

2.370 **JCA:** What about when you are studying, say a Marketing text. Would you not be more careful then?

J: Yes, but sometimes you feel tired, you cannot concentrate. Doing an exam, you will be very careful, but at home you are more relaxed, it's quite different.

JCA: Which is the best sort of reading, do you think? When you are doing the exam, or in your house, more relaxed?

J: Of course, when I am doing the exam, because you concentrate. For instance, at home, when I am reading, sometimes I think of Colombia, I think of my family, or my friend from London phones and I am happy for a long time.

2.380

JCA: So you are easily distracted?

J: Yes, but in the exam, I think hard in order to answer the questions, and I only think about the test, no more.

JCA: When you were answering the test, were you wondering what questions I would ask you or did you not think about that?

J: I did not think about that.

JCA: Do you find it difficult now to talk about what you were thinking about?

2.390 **J:** Yes. My English does not let me express all the things I want to say.

JCA: Say it in Spanish.

J: No, the idea is to learn English, I don't want to say it in Spanish.

JCA: OK. Do you think your performance on this test is typical of your reading ability?

J: Yes.

JCA: Do you think your reading is good?

2.400 **J:** For this test, yes. In English sometimes it takes me longer than other times, but really I can understand the text. But if you have to write, then it takes time, and is more complicated. It is quicker to learn reading than writing or understanding English people.

JCA: Do you think you are a good reader?

J: Why not? Yes.

JCA: In all subjects, or particularly in Marketing?

J: General subjects. Marketing for me is easy to read because I know a lot of words, I know the special words.

JCA: But this was nothing to do with Marketing.

2.410 **J:** But learned journals are worse, difficult words, complicated grammar. For instance, editorials in the newspaper are difficult. Even in Spanish the language they use is very sophisticated, the words, the structure and grammar are very sophisticated because they want to write very elegantly.

JCA: So it is a style problem?

J: Yes, but this is special because journalists try to write elegantly

JCA: Do you do much reading in English?

J: Yes, I try to.

JCA: What sort of things do you read?

2.420 **J:** I try to read magazines, I take seven or eight from the Resource Centre and I try to read articles about sports etc, because I enjoy them, and it's better for me. Because I enjoy, I like to do, I can work.

JCA: What are your main difficulties in reading, do you think? What causes you most difficulty?

J: I don't know. Style, and the kind of topic. For instance in magazines you have a general topic, but when you have a specific magazine or topic, the style and the words are quite difficult. How the text is written.

2.430 **JCA:** Do you have any difficulty reading Spanish?

J: Sometimes, I need to read very carefully. You read and what you read, I don't remember, I need to read again. I read fairly quickly. Not very quickly, but not very slowly.

JCA: Are you worried about your reading when you come to University?

J: Sometimes. There is a relationship between if you are understanding a text and if you are enjoying the text. If you understand, you feel nice and happy. If you don't understand a lot of words, it is very hard, you have to think a lot. When you finish, you feel very tired, because of the thinking and concentration. At university when you finish the week you feel very tired, because you make a very special effort and your mind needs to work more than in Spanish.

2.440

JCA: OK. Good. Thank you very much indeed.

APPENDIX 3

FINAL DISCUSSION WITH J AND P ABOUT THE EXPERIENCE OF TAKING THE TEST.

JCA: Before I give you the answers, P. you said you found that easier, just then?

P: Yes, because I was alone. Many of these questions, I found them. For instance, this question that says "Coronary heart disease" (*number 7*) I found that it was a part here. I want to say that I used to read first the text, but as I can say now, and from every time, I can't say that it is useful every time. For instance, this time it was more useful to start from the questions. If I was alone, I was looking this, I was looking through all the questions very fast. As I did when I was alone, I look until the last question.

3.10

JCA: When you were looking at all the different questions, why were you looking at them?

P: Because I wanted to take the principal idea, the purpose, what you wanted from me, to get an idea of the test.

JCA: Which is actually what J. did when he first sat down. He had a quick look at the text, he looked through it like that, and then went to the questions, had a quick look through the questions, and then began the test. When P. started the test, he didn't look at the questions, he started reading the text, first the first page, then the second page, and then realised that there were five more pages to come, he said, Ah, I'll never finish, so then he

3.20

went back to the questions. OK So it was because I was there, and was asking you to think aloud that you found it more difficult to do. OK. You changed your mind, did you, in some of the questions?

3.30 **P:** Yes, of course. Because I went through all this, got a more general idea of the text, what it is. For instance, I found the bibliography here and I changed my mind in page three. I am not sure that it isn't scientific, because there is all this scientific bibliography, for instance "Epidemiological approaches to the study of cancer and other chronic diseases". It must be scientific.

JCA: Did either of you find it difficult to talk about what you were thinking about?

P: No, I like to talk about it. It was a little amazing, but I didn't have a problem at all, I enjoy it.

3.40 **J:** No, for me it is quite difficult. Especially with English people, because you are making a lot of mistakes, you are thinking of the structure of the sentence, the grammar.

JCA: (*then gives the answers*)

Item 1: P. does not know the phrases given as alternative answers to Item 1). Says he knows the words but not the expressions. Recognises that "due to" is the same, but claims it is not frequently used.

3.50 Item 3: P. claims the wording of Question 3 is unclear ("*Write out the first six words . . .*"), as both he and J. have got it wrong. P. wrote six separate words from Section II, J. referred to the Introduction, instead of Section II. P. claims he does not understand the question and suggests a better wording. "It is the same in Greece, where he has difficulty understanding test items, not the passages themselves."

Item 4: P. did not do, as did not know the words. Thought it would take too much time to read all the Section in order to answer the item, and therefore decided it would be better to go on to other items.

Item 7: P. says it is his fault that he did not go through all the text. Later he went through it and found the heading "Coronary heart disease".

3.60 Item 8: P. feels Item 8 is very easy and students like this type of question. He did it quickly and easily. Other items take much longer. Yet feels it does measure reading ability.

Item 9: P. did not answer. Claims J. did better because he has an Economics background. He himself has a Politics background and hates

diagrams and such. Feels it is important to take account of this background, prefers theoretical studies, and feels people are advantaged if they like diagrams. He reports having spent a lot of time looking at the Table (which he calls a diagram, but did not look at the Figure, as instructed by the item!) He failed to notice the phrase "overleaf" in "Figure 21.3 (overleaf)", but when this was pointed out to him, and he was asked to look at the graph, he still had difficulty interpreting it. Says he does not understand, and JCA explains.

3.70

Item 10: P. did not attempt. JCA gives him time to try it. He takes 2 minutes, but then gives up and JCA stops them.

3.80

P: then offers an opinion on testing in general, claims he hates exams because they are not representative of his ability, and they make him anxious. However, he then says that both the Wrestling test and the Smoking test are different and not difficult. They are like a puzzle, or a game, and are interesting. Claims teacher can discover his knowledge from the performance, and he does not feel depressed. Recognises there are no consequences on passing or failing the test, yet admits to some stress on taking the Smoking test.

BIBLIOGRAPHY

- Adams-Smith, D. (1981) "Levels of Questioning: Teaching Creative Thinking through ESP" *Forum*, Vol 19, no 1 pp 15-21
- Alderson, J Charles (1987) "Testing Reading Comprehension; The Notion of Hierarchically Ordered Skills" Plenary address, First International Conference on Language Testing, Tsukuba University, Tokyo, Japan, March 1987.
- Alderson, J. Charles and Grant Henning (1987) "Testing Reading" Paper given at the Conference on Developments in Language Testing Research, University of Reading, UK, April, 1987.
- Alderson, J. Charles, Grant Henning and Yasmeeen Lukmani (1987) "Levels of Understanding in Reading Comprehension Tests" Paper presented at the Ninth Annual Language Testing Research Colloquium, April 1987, Miami, USA.
- Alderson, J. Charles and Yasmeeen Lukmani (1989) "Cognition and Reading: Cognitive Levels as Embodied in Test Questions" *Reading in a Foreign Language*, Vol 5, no 2 pp 253-270
- Alderson, J. Charles and A.H. Urquhart (1984) *Reading in a Foreign Language*. Longman.

- Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H. and Krathwohl, D.R. (1956) *Taxonomy of Educational Objectives. Handbook I: Cognitive domain*. New York: Longmans, Green
- Clymer, T. (1968) "What is reading?: some current concepts" Reprinted in Melnik, A and J Merritt (1972) *Reading Today and Tomorrow* University of London Press, pp 48-66.
- Criper, C. and A. Davies (eds) (1988) *ELTS Validation Project Report*. English Language Testing Service Research Report Vol 1(i). The British Council and The University of Cambridge Local Examinations Syndicate.
- Gray, (1960) "The Major Aspects of Reading" in Robinson, H (ed) *Sequential Development of Reading Abilities* Supplementary Educational Monographs, no 90. Chicago University Press, pp 8-24.
- Lunzer, E. Waite, M. and Dolan, T. (1979) "Comprehension and Comprehension Tests" in Lunzer, E. and Gardner, K. (eds) *The Effective Use of Reading* Heinemann Educational Books, pp 37-71.
- Munby, J. L. (1978) *Communicative Syllabus Design* Cambridge University Press.
- Oller, J. W. (ed) (1983) *Issues in Language Testing Research* Rowley, Mass: Newbury House.
- Seddon, G. M. (1978) The Properties of Bloom's Taxonomy of Educational Objectives for the Cognitive Domain. *Review of Educational Research*, Vol 48, no 2, pp 303-323.
- Weir, C. J. (1983) *Identifying the Language Problems of Overseas Students in Tertiary Education in the UK*. Unpublished PhD thesis, Institute of Education, University of London.