Reading Comprehension Assessment: From Text Perspectives

MIYOKO KOBAYASHI
Kanda University of International Studies

This paper investigates the nature of reading comprehension questions. Very few studies have so far examined comprehension questions in relation to text features. Kintsch and Yarbrough (1982) and Shohamy and Inbar (1991) are among the few studies, and their results suggest that there is an interaction between text features and the focus of questions. The present study builds on these findings and examines how Meyer's (1975, 1985) model of content structure analysis can help identify what exactly reading comprehension questions try to measure. The paper then proposes a framework for characterizing comprehension questions, and examines the characteristics of 40 questions in terms of their interrelationship of item types and item statistics based on the test results of 227 Japanese university students. Finally, implications for assessment practice are discussed.

Introduction

It is now widely recognized that reading process is an interaction between the reader and the text (e.g. Carrell and Eisterhold 1983), and it is important to take both of them into account in the discussion of reading comprehension. It has even been suggested that a text only has meaning potential to the extent that every reader brings his or her unique background knowledge when reading a text. However, in the assessment of reading, such infinite potential in meaning is problematic as the testers need to have 'correct' answers to assess learners' comprehension. In this connection, the distinction between 'comprehensions' and 'interpretations' (Urquhart 1987) would

be useful. The latter relies on the reader's background knowledge, and therefore goes beyond the text. By contrast, the former relates to the understanding of the main message contained in the text. Even though 'comprehensions' could involve some degree of inference, the inference should logically be deducted from the information in the text (so-called 'reading between the lines', not 'beyond the lines'). Therefore, it would be worth investigating what is to be understood in a text in the first place to understand the nature of reading comprehension questions.

In this paper, I shall first discuss reading sub-skills hierarchy. Then I shall introduce some approaches to text analysis and examine a model of content structure analysis to understand what is involved in reading comprehension questions. The characteristics of question types are further examined by analyzing the comprehension questions prepared for another larger-scale study (Kobayashi 1995, 2002).

1. Reading Sub-Skills Hierarchy

A number of reading sub-skills taxonomies have been proposed (e.g. Bloom 1956; Grabe 1991; Lunzer & Gardner 1979; Munby 1978), and interesting research and discussions have been under way in relation to sub-skills identification and skills hierarchy (e.g. Alderson, 1988, 1990a, 1990b, 1993; Alderson & Lukmani, 1989; Lee & Musumeci 1988; Lumley, 1993, 1995; Matthews, 1990; Rost 1993; and Weir *et al.*, 1990). A notable distinction may be between a literal understanding of a text and an understanding of implicit ideas. Another distinction could be between understanding details and understanding the main ideas of a text (Alderson 2000).

Following the premise that reading ability can be divided into sub-skills, it has long been discussed that these sub-skills can be ranked hierarchically from lower-order to higher-order skills according to the nature of the reading processes involved. Lower-order skills are supposed to require local level

understanding such as word recognition or literal understanding whereas higher-order skills require higher levels of cognition such as an ability to synthesize, infer, and evaluate (see Buck, 1990). There seems to be agreement in defining a process as 'higher' when it involves more of the reader, and 'lower' when it is mainly text-based. It is generally thought that lower-order skills are easier than higher-order skills (see Alderson 1979, 1980), though discussion on the issue has been inconclusive. Research evidence further suggests that so-called 'inference' questions are poor discriminators (Buck 1990; Perkins and Brutten 1988).

The discussion of sub-skills taxonomy and hierarchy leads to further questions: are these sub-skills one-dimensional? Do they interact with other factors such as the reader's familiarity or text organization? For example, it seems quite obvious that 'low-level' understanding of a word can be difficult if the reader is not familiar with the word concerned. On the other hand, drawing inferences can be easy when the reader has sufficient background knowledge about the topic. Similarly, identifying main ideas can be easy if the text is well-organized and clearly indicates them. In other words, item difficulty may also depend on factors other than types of reading skills. Weir *et al.* (1990) rightly argue that we cannot compare the difficulty of two separate items simply by their skill levels because different sets of skills operate independently of each other.

Moreover, there seems to be some confusion in terminology. The distinction between higher- and lower-order skills can be interpreted in different ways. Kintsch and Yarbrough (1982) is one of the few studies that examined the relationship between test format and reading constructs. However, their view that short answer questions could measure global understanding while cloze items only touched upon local understanding seems rather simplistic. Similarly, in the area of listening comprehension, Shohamy and Inbar (1991) investigated the effects of question types (trivial vs. global) and text features on

comprehension. The results showed that the effects of 'orality' on comprehension varied more in trivial questions. However, they seem to confuse two different aspects of questions: 'micro- vs macro-level,' on one hand, and 'literal vs. inference,' on the other. It is necessary to understand the exact meaning of these terms before interpreting the results and drawing conclusions.

The work of Pearson and Johnson (1978) may offer a useful insight into the discussion of sub-skills hierarchy. They categorize reading comprehension questions into three types in the light of the interactive role between the reader and the text: textually explicit; textually implicit; and 'scriptally' implicit. The first type of question is related to information explicitly stated in a single sentence in the text. The reader's task is to identify the information and understand it literally. Textually implicit questions are concerned with information stated in different sentences in the text and the reader has to integrate the separate pieces of information to answer correctly. The third type, 'scriptally' implicit questions, requires the reader to combine information available in the text with his or her prior knowledge. This type of question can be called 'inference' questions in conventional taxonomies. Among the three question types, the reader's role reaches its maximum with the third type. This three-level classification of comprehension questions helps clarify the nature of questions in terms of the relationship between the reader and the information in the text.

The question is still open as to whether sub-skills exist or whether they are hierarchically ordered. However, as Urquhart and Weir (1989: 93) suggest, a list of sub-skills are 'useful tools for the development of both teaching materials and tests. In spite of the doubts that have been raised, we shall continue to make use of the taxonomies.'

In the next section I shall turn to what text analysis can offer for our understanding of reading comprehension, especially with regard to what exactly we understand in a text.

2. What do we understand?

2.1. Approaches to text analysis

There have been various attempts to break texts down into segments to account for the process of comprehension. One approach is to try to identify idea units in relation to other ideas in the text. The most notable examples of this approach are probably Kintsch and his associates (1973, 1975, 1977, 1978) and Meyer (1975, 1985). Kintsch *et al.* (1975: 196) use the term 'proposition' for an idea unit and define 'propositions' as follows:

> "... the basic units of meaning are *propositions*. Propositions are n-tuples of *word concepts*, one of which serves as a *predicator*, and the remaining ones as *arguments*, each fulfilling a unique semantic role. The predicator specifies a relationship among the arguments of a proposition." (emphasis in original)

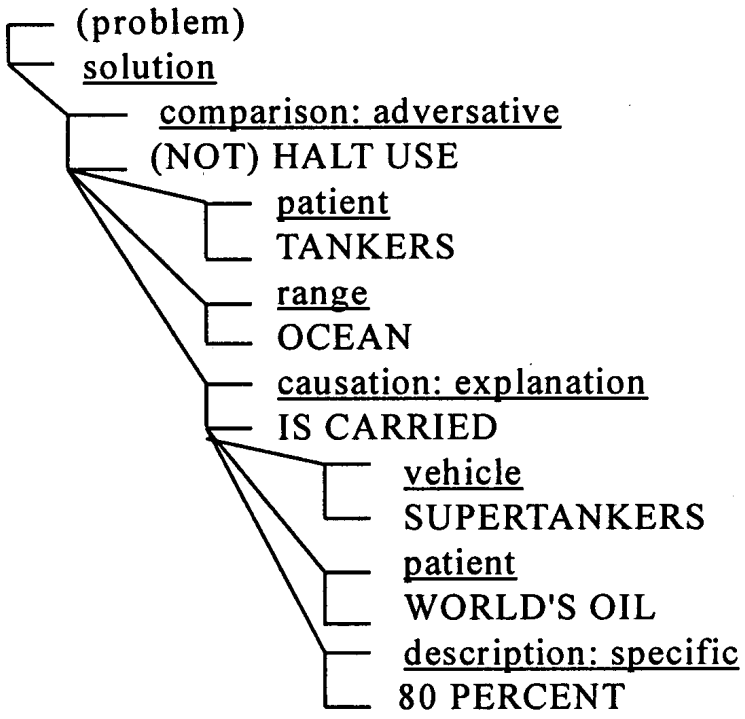
Kintsch's propositional analysis of texts has been used by many other researchers, especially in studies on summary writing and recall protocols (e.g. Britton *et al.*, 1982; Dixon *et al.*, 1984; Duffy *et al.*, 1989; Vipond, 1980). Kintsch and van Dijk (1978) further developed the concept of macro-structure, a global structure of a text, on the basis of the hierarchical relationship of propositions. The problem of this scheme, however, is that repetition and intuition are the basis of constructing the hierarchy; there are no other clear guiding principles. This is where Meyer's (1975, 1985) content structure analysis seems more helpful.

Meyer's analysis is called 'semantic content structure analysis'. This system has been developed on the basis of Fillmore's (1968) case grammar and Grimes's (1975) semantic grammar of propositions. According to Meyer, each idea unit is assigned a role in relation to others. The roles include: 'Agent', 'Instrument', 'Force', 'Vehicle', 'Patient', 'Benefactive', 'Latter', 'Former', and 'Range'. At the same time, rhetorical relationships between ideas are considered. The relationships

include: 'alternative', 'response', 'collection', 'attribution', 'time', and 'specific'. The 'collection' relation, for example, is a list of elements related to each other in some way or other. The 'specific' relation is what Meyer calls 'hypotactic' relation because one idea gives more specific information about another. Based on these roles and relationships, idea units are then arranged in a hierarchical order. The following figure illustrates how ideas are represented in the content structure.

Figure 1. Example of content structure diagram

Ex. The solution to the problem is not to halt the use of tankers on the ocean since about 80% of the world's oil supply is carried by supertankers.



(Meyer 1985: 286-287)

This hierarchical structure then constitutes a larger rhetorical structure of a text. The rhetorical relation which appears at the top of the hierarchy is called the top-level rhetorical structure and this characterizes the structure of the whole text. The top-level rhetorical structure is identified as one of the following: 'collection', 'causation', 'response', 'description' and 'comparison' (Meyer later renamed 'response', calling it

'problem-solution'). These five types of top-level relationships are thought to represent patterns in the way we think (Meyer 1985: 20).

Thus, the Meyer analysis attempts to incorporate the relationships between ideas, whether explicit or implicit. This is particularly useful when considering comprehension; it clarifies what we understand and how information is organized in our mind. The value of the rhetorical predicate as a unit for text analysis has been appreciated by a great number of other researchers, and Meyer's model has been applied in the interpretation of recall protocols and summary writing. For example, Golden *et al.* (1988: 140) quote Mann and Thompson (1986) arguing:

> "... rhetorical predicates contribute to the connectivity, coherence, and function of texts. The connecting of these related parts of a text into a coherent whole occurs when the reader interprets the text a writer has intentionally written; thus rhetorical predicates are central to this interpretive process."

Empirical studies conducted by Meyer (1975, 1987), Meyer and her associates (e.g. Meyer *et al.* 1980; Meyer *et al.* 1993; Meyer, Brandt & Bluth, 1980; Meyer & Freedle 1984; Meyer & McConkie, 1973; Meyer & Rice 1984), and many other researchers have shown that ideas that come higher in the content structure are better recalled after reading or more likely to appear in summaries. In other words, the position of ideas in the content structure hierarchy is a good predictor of recall.

In conventional reading sub-skills taxonomies, the terms 'main ideas' or 'supporting ideas' have never been clearly defined. Meyer's content structure analysis can make a contribution here. Ideas are hierarchically arranged in the content structure diagram and those high in the hierarchy can be defined as main ideas on the one hand while those that are lower in the hierarchy rank merely as supporting details on the other.

Based on her research studies, Meyer (1975) posits that the content structure will be useful for constructing comprehension questions because it allows us to identify the relative importance of a particular piece of information in a text. Meyer (1975: 182) says:

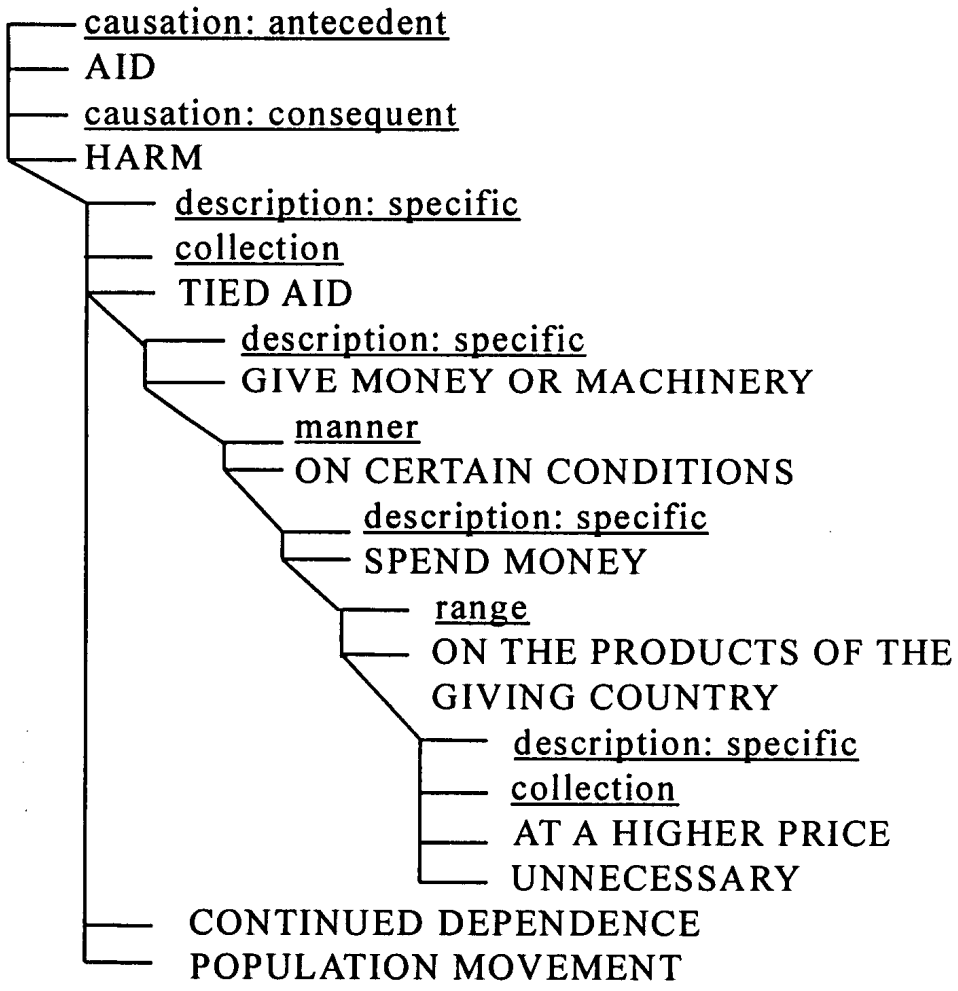
> "... the content structures ... appear to have great potential for use in the designing and classifying of questions from passages, as a guide for writing questions ... [Q]uestions could be classified according to the location of information required to answer them in the content structure."

However, Meyer herself did not investigate comprehension questions. In the next section I shall illustrate how the content structure can be applied to the analysis of actual reading comprehension questions and how such analysis can shed light on our understanding of the nature of comprehension questions.

2.2. An Examination of Reading Comprehension Questions in the Light of Content Structure Analysis

Figure 2 below is a part of the content structure analysis of an 'International Aid' text (see Appendix). In this content structure diagram, the top-level organization is 'causation': 'aid causes harm', and it is represented at the top left corner. 'Tied aid', 'continued dependence' and 'population movement' are given as examples of 'aid causing harm'. The description about 'tied aid' is given under the entry of 'TIED AID' at several subordinate levels.

Figure 2. Content structure of International Aid Text



(adapted from Kobayashi 1995)

In constructing reading comprehension questions about the idea of 'tied aid', several alternatives are possible such as:

1. What is tied aid?
2. What do you call a type of aid which is given on certain conditions?
3. Why is tied aid a problem?
4. Give an example which illustrates that the aid given to Third World countries has an undesirable effect.

Let us consider what kind of information is needed to answer these questions in the light of the content structure diagram. The necessary information to answer the first question is contained in the first reference to 'TIED AID' and the

elaboration which follows. It could be argued that the first part of the description (GIVE MONEY OR MACHINERY, ON CERTAIN CONDITION) is sufficient as an answer to the question. Alternatively, the whole section may be required for a really full answer. Therefore it is hard to give a precise definition of the amount of information which is necessary. This suggests that it may also be necessaray to take into account the amount of information required, in addition to the information's height in the istructure, when the nature of questions is considered. In spite of this lack of precision, however, it is relatively easy to find the information required in the content structure. Furthermore, since these ideas are at the third level in the hierarchy or lower, it can be suspected that the ideas are not particularly important.

The situation is fairly similar in the second question. One difference is that the reader has to find the description first and then the relevant term which corresponds to the description. This process may be more troublesome because it requires the reader to start from details, at lower levels, and then go back to a heading at a higher level. Another difference is the length of answer required: only two words are required for the second question while a fairly long explanation is required for the first. This also relates to the amount of information to be processed: to answer the second question, the reader has only to find a part defining 'tied aid', without having to fully understand it. This will certainly reduce the load on the test-taker.

All the same, the first two questions are more or less similar in terms of the amount of context needed. The reader has only to go to a relevant section of the text and reproduce necessary information as an answer. In a sense, this is similar to a cued recall; the only difference - which is a big one - is that the reader is allowed to refer back to the text in the case of reading comprehension tests. In such cases, identification of necessary ideas in the content structure is straightforward, and the ease of answering the question may be related to the height in the

hierarchy of the content structure as in recall.

The third question is more complex. In order to be able to answer it, the reader must not only understand the description about 'tied aid', but he or she must also grasp the overall structure of the text to understand that 'tied aid' is an example of 'aid causing harm'. This task requires the reader to read a fairly large amount of text, to understand how the ideas interact with each other in the content structure and to integrate these ideas to produce an appropriate answer. In other words, the amount of cognitive load required here is greater than in the first two questions. Therefore, the third question is different in nature from the first two, and is expected to be more difficult to answer as a result. The fourth question is similar to the third in that it requires a wider range of context than the first two and a grasp of the top-level structure. However, it may be easier than the third for two reasons: first, the amount of information the reader has to produce as an answer is much smaller (this will make a big difference if the answer is expected in the target language) and, secondly, the reader can choose one example, which he or she is more confident about, out of the three given in the text.

The above example clearly illustrates how content structure analysis can help to identify what is involved in particular comprehension questions. It has also shown that, even if a particular idea is identified as a main focus of a question, there are various ways of framing comprehension questions to explore this idea, varying the difficulty levels of questions. It is not sufficient to locate the position of the idea in the content structure to understand exactly what is involved in the question. Quite frequently, more than one idea is necessary to answer a question. Therefore, it is not easy to classify questions 'according to the location of information needed to answer them in the content structure,' as Meyer claims (1975:182), because the picture is far more complex in reality. This may be due to the complex nature of reading comprehension questions.

Before moving to the next section, it must be pointed out that there are some difficulties in applying the Meyer model of content structure to reading comprehension assessment. One is that it is time-consuming to construct the diagram in the first place. Meyer suggests that implicit ideas and relationships should appear in the diagram, but it is not clear how exactly they should be expressed. There is no rigid rule with regard to the exact wording or the amount of information which should be represented in the diagram.

Another problem is that it is even more difficult to construct a diagram in less well-organized texts, where there is no clear top-level structure; several sets of information appear independently and they do not link with each other. Of course, this may be why the text is categorized as 'loosely organized' and the difficulty may therefore be inevitable, but it raises the question how to apply Meyer's analysis to handle such texts, which are not uncommon in real life. One implication of this could be that the nature of comprehension questions based on loosely-organized texts are more likely to be difficult to identify.

Thirdly, the content structure has a limitation in dealing with a wide aspect of reading comprehension. Earlier in this paper I noted that the assessment of reading should limit itself to 'comprehensions', excluding the reader's idiosyncratic interpretations, and introduced Pearson and Johnson's three-level categorization of reading comprehension questions: textually explicit, textually implicit, and scriptally implicit. The first of these is handled with relative ease. The second kind of reading is not so straightforward, but, as the illustration above shows, the content structure at least helps us to understand what is involved in the questions. However, the third kind, i.e. scriptally implicit, seems to present a problem. The content structure is claimed to incorporate implicit relationships between ideas, but they do not go beyond rhetorical relationships. Inference questions require the reader to read

between the lines and respond in relation to common sense, value system, cultural reference, etc.

Let us go back to the text discussed above. The diagram shows that 'Aid causes harm.' First of all, we need to know that 'harm' is an undesirable state of affairs. Then we can deduce that it is not good when something causes harm. With our common sense, we can see that the author is not just describing something like a scientific experiment in which one thing causes another, but is trying to convey a message that the current state of aid is unsatisfactory. This is all implicit; the author does not use words such as 'bad' or 'undesirable' or 'unsuccessful.' Such values associated with the word 'harm' do not appear in the content structure. Therefore, the content structure does not give a clear explanation about what is needed to answer a question such as: 'Does the writer of the passage think that giving aid is generally successful? Give reasons.' This limitation is, in a sense, inevitable because Meyer started her text analysis on the basis that comprehension equaled to reproducing the text by recall. Answering reading comprehension questions is evidently quite different from recall.

Nevertheless, it is clear that, as the above illustration shows, content structure analysis can help identify the range of information required to answer comprehension questions. It is now evident that the first two questions examined above are very different from the last two in their nature. Language teachers and testers may already know this by intuition and experience, but they would find it more illuminating to have a diagram like this at their side and to know exactly what kind of questions they are producing and what they are asking their students to do by setting particular comprehension questions.

2.3 Framework for categorizing reading comprehension questions

Meyer's content structure analysis has implications for the issue of sub-skills hierarchy discussed above. The above analy-

sis suggests that the amount of information required to process to answer a question and the height (i.e. the importance) of the information in the hierarchy seem somewhat independent of each other. For example, understanding the main ideas does not necessarily require global understanding if the ideas are presented explicitly, possibly with clear signals. In other words, weighing up the importance of ideas and dealing with a certain amount of information are two separate processes, though possibly interacting with each other. Therefore it would be helpful if reading comprehension questions are categorized in two textual aspects: amount of information and the importance of the information. Incorporating Pearson and Johnson's (1978) three-level distinction mentioned above, the following framework could be proposed to capture three different dimensions of reading comprehension questions, each having three-levels:

a) Reader's involvement
- literal understanding (textually explicit)
- integration (textually implicit)
- inference (scriptally implicit)

b) Amount of information
- global (beyond a paragraph)
- medium (beyond a sentence)
- local (within a sentence)

c) Importance of information
- main ideas (top level)
- supporting ideas (medium level)
- details (low level)

On the basis of this framework, reading comprehension questions will further be examined in the next section.

3. Characterizing comprehension questions according to the question type framework

In this part of the paper, I shall report the results of an empirical investigation into the characteristics of different types of reading comprehension questions in the light of the framework proposed above. I shall examine a set of actual comprehension questions, which were prepared for another larger-scale study which investigated the effects of text organization and test format on comprehension test performance (Kobayashi 1995, 2002).

3.1 Analysis

In total, 40 short-answer reading comprehension questions, five each for eight texts of 350 ~ 380 words, were analyzed. The analysis included: examining the inter-relationship between the three question type categories; and investigating the characteristics of comprehension questions in terms of item statistics. The item statistics were obtained from the test results of the total of 227 Japanese university students. They were all native speakers of Japanese, with similar educational and social backgrounds, based in one of eight universities in Japan. Roughly speaking, two thirds were female students and one third male. The majority were either first or second year students, taking English as a required subject. They had had six years' English learning at secondary schools. Their English proficiency level was roughly from lower-intermediate to intermediate.

After all the questions were prepared, their question types were identified in two phases by expert judges, who were mostly educated native speakers of English. Most of them were MA holders in applied linguistics and were currently engaged in EFL teaching, materials development or testing consultancy. When non-native speakers were involved, they were invariably fluent English speakers and holders of MA degrees in applied linguistics or education. In the first phase of analysis, five

experts were invited to identify the questions according to characteristics based on the three categories described above. After some modifications of questions taking account of the problems identified, a further 16 experts were asked to make their judgments. 95% of the items achieved agreement among the majority of the judges with regard to the question types. The remaining 5 % which did not attain the majority agreement (two items in the category of reader involvement) were excluded from the analysis.

3.2. Results

3.2.1. Question Type Categories

The following tables (1-3) show the inter-relationship between the three question type categories, and the data demonstrate interesting trends.

Table 1 shows that literal questions tend to require smaller amounts of context than integrative questions. There are no examples of literal questions with the context beyond a paragraph level whereas in integrative questions or inference questions there are no examples below the sentence level. Table 2 suggests that there is a slight tendency for questions asking about more important ideas to require a larger amount of context, but the relationship between the two categories is not so clear as in Table 1. Table 3 shows a less straightforward relationship between the reader's involvement and the importance of information. The great majority of integrative questions are related to important ideas whereas literal questions tend to be related to less important ideas. Interestingly enough, no discernible pattern emerged in inference questions in terms of the importance of ideas.

Tables 1-3: Relationship between question type categories

Table 1: Reader's involvement and the required amount of context

	Local (n=8)	Medium (n=24)	Global (n=8)
Literal (n=15)	8	7	0
Integrative (n=16)	0	10	6
Inference (n=7)	0	5	2

Table 2: Required amount of context and the importance of information

	Low (n=13)	Middle (n=14)	Top (n=13)
Local (n=8)	5	3	0
Medium (n=24)	8	11	5
Global (n=8)	0	0	8

Table 3: Reader's involvement and the importance of information

	Low (n=8)	Middle (n=24)	Top (n=8)
Literal (n=15)	8	7	0
Integrative (n=16)	2	3	11
Inference (n=7)	3	2	2

The above findings seem to contribute to the discussion of 'higher- vs lower-order skills' or 'macro-vs micro-level questions', which have been used in reading comprehension circles without being clearly defined. It seems to be widely believed that 'lower-order' items are literal questions requiring a small amount of context and concerned with details, whereas 'higher-order' items are integrative or inference questions requiring a larger amount of context and concerned with important ideas. However, as the data show, these characteristics are somewhat independent of each other and not all inference questions, for example, require a large amount of context or are related to important ideas. This suggests that it is more beneficial to take a multidimensional view like the one presented here than a one-dimensional dichotomy such as 'higher-/lower-order' or 'micro-/macro-level' distinctions.

3.2.2. Item Statistics

Furthermore, the item statistics suggest that different types of questions behave differently as shown in Table 4 below. FV stands for 'facility value,' as an index of item difficulty, and IT stands for 'item-total correlation', as an index of item discrimination. Item discrimination is an ability of an item to distinguish between more able and less able test-takers. High item discrimination suggests that the item is effective in reflecting learner's overall ability.

Table 4: Item Statistics according to Question Types

Type		No. of items	FV	IT
1	Literal	(n=15)	.57	.38
	Integrative	(n=16)	.41	.50
	Inference	(n=7)	.32	.40
2	Local	(n=8)	.65	.34
	Medium	(n=24)	.40	.44
	Global	(n=8)	.42	.48
3	Low	(n=13)	.51	.40
	Middle	(n=14)	.43	.37
	Top	(n=13)	.43	.52

1: reader's involvement; 2. amount of information required;
3: importance of information

The data in the table show that literal questions, items requiring local understanding (within a sentence), and items asking about less important information tended to be easier, and that integrative questions, items requiring global understanding (beyond a paragraph), and items asking about important ideas achieved higher item discrimination.

Analyses of variance (ANOVA) were conducted to see whether there was any statistical significance in the above trends, and the results supported the following:

- Reading comprehension questions tended to be easy

  a) when they were related to literal understanding (F (2, 35) = 7.84*, p< .005), and

b) when they required a sentence-level context (F (2, 37) = 8.34*, p< .005).

- Reading comprehension questions tended to be more reliable and better reflect learners' language proficiency

a) when they required integration of ideas (F (2, 35) = 3.71*, p< .05), and
b) when they asked about important ideas (F (2, 37) = 5.16*, p< .05).

The above findings seem to illuminate the characteristics of reading comprehension questions and their effects on test performance. In particular, the identification of question types which are more efficient in discriminating between learners (i.e. integrative questions and questions about important ideas) will have useful implications for reading assessment.

4. Summary and Discussion

This paper has examined the nature of reading comprehension questions. Owing to its limitations regarding the small numbers of questions examined in the study, the results should not be over-generalized, but they have presented interesting implications for the assessment of reading comprehension.

First, comprehension questions were examined in the light of Meyer's model of content structure analysis. The analysis helped clarify what exactly was involved in answering questions in terms of the importance of information and the amount of information to process to answer the question. As fully discussed in 2.2, the analysis identified the complex nature of reading comprehension questions. Thus, Meyer's content structure diagram proved to be a useful tool to distinguish differences in the nature of different comprehension questions, despite its limitations.

Secondly, comprehension questions were analyzed and characterized according to the three question type categories.

The investigation has revealed interesting interrelationships between the question types, and has suggested that the characteristics of comprehension questions are far more complex than can be captured in a one-dimensional dichotomy such as a higher-/lower-order or macro-/micro- level distinction. This study was inspired by Kintsch and Yarbrough (1982), who suggested that short answer questions were more sensitive to global understanding than cloze tests. They specifically used short answer questions relating to main ideas such as 'What is this passage about?', but they did not doubt that there were a variety of types of questions or how it could be ensured that the questions were about main ideas.

A further examination of item statistics has revealed that test performance could vary according to the types of questions. For example, comprehension questions tended to be easier when they required literal understanding or a smaller amount of context. On the other hand, questions tended to be more reliable and reflect the learner's language ability better when they were related to important ideas in the text or require integration of ideas. Some of the findings were statistically significant. This suggests that we need to pay more attention to these different characteristics of comprehension questions because different types of questions seem to touch upon different aspects of reading comprehension and to vary in their test efficiency.

The data suggest that the three-dimensional framework proposed here is useful for characterizing reading comprehension questions. The difference between this framework and existing taxonomies of reading sub-skills (e.g. Munby 1978) is that it takes a dual perspective of reading: the reader-oriented and text-oriented, and the categories are hierarchically ordered within each domain (see Skehan, 1988, for criticism of the Munby list). The results of expert judgment seem to suggest that the framework is less problematic to apply than existing sub-skills taxonomies (e.g. Criper and Davies 1988). Both

language testers and classroom teachers will benefit by employing such frameworks in item construction and item analysis. A fuller understanding of question items by examining them against this kind of framework will be particularly useful for ensuring the extent of coverage of reading skills in a particular test battery.

Of course the framework proposed in this study is not without problems. The first category, according to the reader involvement, seemed to be difficult to apply to actual questions. In particular, the experts expressed uncertainty about the distinction between literal and integrative questions. It may be that item characteristics form in a continuum rather than dividing into a clear-cut set of categories (e.g. Rosenshine 1980). At one end, there are items which are clearly text-based and require little or no integration on the part of the reader, and at the other, the reader's involvement is at its maximum, as in inference questions. For example, identifying a specific piece of information in the text requires very little integration. Further up the continuum, if a question is related to two pieces of information, the nature of the questions can vary and this is where the experts in this study had difficulty. When the two pieces of information are closely adjacent to each other, they can easily be identified as if they form one piece of information. On the other hand, if they appear in different parts of the text, the amount of work the reader has to undertake becomes greater, even if the information itself is explicit. Furthermore, a question becomes even more demanding when the reader has to read a number of sentences to grasp an idea. This certainly requires more integration on the part of the reader. Furthermore such complexity of question items is multiplied by the difficulty of defining or counting an 'idea'.

The third category, i.e. the importance of information, has also presented a problem. This ought to be fairly straightforward as long as the decision is based on the location and/or height of an idea unit in the content structure diagram.

However, the degree of importance perceived by the experts in this study varied considerably. This may be explained by a number of factors. For example, the experts' previous experience and schemata may vary and this may have affected their judgment as to what is important. Alternatively, the difficulty may be related to the difference between recall and comprehension questions. When dealing with recall protocols, idea units in the original text are the basis for analysis. Therefore it is straightforward to identify the location of particular idea units in the content structure diagram (see Meyer 1985: 289-296). This automatically determines their 'height' in the hierarchy, i.e. importance. By contrast, reading comprehension questions are more complex. As shown in the analysis illustrated above, they are often related to a wider range of information: such as two top-level ideas or a set of ideas ranging from high to low in the hierarchy. In other words, there is no one to one correspondence between the information required for answering a question and each idea unit in the hierarchy. This variability involved in reading comprehension questions seems to cause difficulty in identifying the importance of ideas required to answer the question.

The findings presented here are particularly significant for item-writers and language teachers. The data suggest that even though short-answer questions are one of the most common formats in reading comprehension tests, they involve a great deal of complexity. In order to assess reading ability accurately, we need to know for sure what is involved in the questions we construct. It is hoped that not only language testers but all language teachers who use comprehension questions will be aware of this variability across different types of questions.

Furthermore, it is perhaps worth mentioning, although the results were not presented here, that, in examining the distribution of different types of questions which were produced for the eight texts, there seemed to be a close relationship between text type and question type (see

Kobayashi 1995). For example, tightly-organized texts tended to produce more questions on main ideas than less well-organized texts. At least, it was easier to generate a variety of questions when texts were tightly-organized. If texts were loosely-organized, on the other hand, questions tended to focus on details or literal understanding. As the data presented above suggest that integrative questions or questions about important ideas are more efficient in discriminating between learners, it would perhaps be more desirable to use more tightly-organized texts in reading comprehension tests to maximize test efficiency.

One of the major limitations of this study was the small number of questions examined. It was difficult to produce more than five good questions about a text of 350 ~ 380 words. Since the data presented interesting implications and this is an unexplored area, it certainly merits further investigation. In order to do so, more texts and/or longer texts would be desirable.

The three-dimensional categorization of comprehension questions proposed in this study would benefit from empirical evidence in future research. One possible approach is to conduct factor analysis to see whether any common factors characterize different question types. Another approach could be to collect test-takers' introspective data to investigate their thought processes in answering questions.

Conclusion

It is often assumed that reading ability is relatively easy to assess, in comparison with other language skills. However, the analysis of this paper suggests that even writing simple short-answer questions of reading comprehension, one of the common test formats, could involve a range of factors and require a great amount of attention to ensure that the test measures what is intended to measure. Test results are often used to make important decisions, whether for educational or

research purposes. It is not an exaggeration to suggest that such decisions often determine life paths, as in university entrance examinations. Therefore it is vital for language testers, or anyone involved in assessment, to pay greater attention to the validity of their tests.

References

Alderson, J.C. 1979. The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13.2: 219-228.

Alderson, J.C. 1980. Native and non-native performance on cloze tests. *Language Learning*, 30.2: 219-223.

Alderson, J.C. 1988. Testing reading comprehension skills. Paper given at the Sixth Colloquium on Research in Reading in a Second Language. TESOL, Chicago. March 1988.

Alderson, J.C. 1990a. Testing reading comprehension skills (Part 1). *Reading in a Foreign Language*, 6.2: 425-438.

Alderson, J.C. 1990b. Testing reading comprehension skills (Part 2). *Reading in a Foreign Language*, 7.2: 465-503.

Alderson, J.C. 1993. Judgments in language testing. In Douglas, D. & Chapelle, C. (eds.) *A New Decade of Language Testing Research.* Virginia: TESOL., 46-57.

Alderson, J.C. 2000. *Assessing Reading.* Cambridge: Cambridge University Press.

Alderson, J.C. & Lukmani, Y. 1989. Cognition and reading: cognitive levels as embodied in test questions. *Reading in a Foreign Language*, 5.2: 253-270.

Bloom, B. 1956. *Taxonomy of Educational Objectives Book 1: Cognitive Domain.* London: Longman.

Britton, B.K., Glynn, S., Meyer, B.J.F. & Penland, M.J. 1982. Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74.1: 51-61.

Buck, G. 1990. The Testing of Second Language Listening Com-

prehension. Unpublished Ph.D. Thesis. University of Lancaster.

Carrell, P.L. & Eisterhold, J.C. 1983. Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17.4: 553-573.

Criper, C. & Davies, A. 1988. *ELTS Validation Project Report 1 (i)*. London: The British Council & University of Cambridge Local Examinations Syndicate.

Dixon, R.A., Hultsch, D.F., Sinon, E.W. & van Eye, A. 1984. Verbal ability and text structure effects on adult age differences in text recall. *Journal of Verbal Learning and Verbal Behavior*, 23.5: 569-578.

Duffy, T.M., Higgins, L., Mehlenbacher, B., Cochran, C., Wallace, D., Hill. C., Haugen, D., McCaffrey, M., Burnett, R., Sloane, S., & Smith, S. 1989. Models for the design of instructional text. *Reading Research Quarterly*, 24.4: 434-457.

Fillmore, C.J. 1968. The case for case. In Bach, E. & Harms, R. (eds.) *Universals in linguistic theory.* New York: Holt, Rinehart, and Winston, 1-81.

Golden, J., Haslett, B. & Gauntt, H. 1988. Structure and content in eighth-graders' summary essays. *Discourse Processes*, 11.2: 139-162.

Grabe, W. 1991. Current developments in second language reading research. *TESOL Quarterly*, 25.3: 375-406.

Grimes, J.E. 1975. *The thread of discourse.* The Hague: Mouton.

Kintsch, W.1977. On comprehending stories. In Just & Carpenter (eds.) *Cognitive processes in comprehension.* Hillsdale, N.J.: Lawrence Erlbaum, 33-62.

Kintsch, W. & Keenan, J. 1973. Reading rate as a function of number of propositions in the base structure of sentences. *Cognitive Psychology,,* 6: 257-274.

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G. & Keenan, J.M. 1975. Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 14.2: 196-214.

Kintsch, W. & van Dijk, T.A. 1978. Toward a model of text comprehension and production. *Psychological Review*, 85.5: 363-394.

Kintsch, W. & Yarbrough, J.C. 1982. Role of rhetorical structure in text comprehension. *Journal of Educational Psychology*, 74.6: 828-834.

Kobayashi, M. 1995. Effects of text organisation and response format on reading comprehension test performance. Unpublished Ph.D Thesis. Thames Valley University.

Kobayashi, M. 2002. Method effects on reading comprehension test performance: text organisation and response format. *Language Testing*, 19.2: 191-218.

Lee, J.F. & Musumeci, D. 1988. On hierarchies of reading skills and text types. *Modern Language Journal*, 72.2: 173-187.

Lumley, T. 1993. The notion of subskills in reading comprehension tests: an EAP example. *Language Testing*, 10.3: 211-234.

Lumley, T. 1995. Responses and replies. *Language Testing*, 12.1: 121-130.

Lunzer, E. & Gardner, K. (eds.) 1979. *The Effective Use of Reading*. London: Heinemann.

Mann, W.C. & Thompson, S.A. 1986. Relational propositions in discourse. *Discourse Processes*, 9.1: 57-90.

Matthews, M. 1990. Skill taxonomies and problems for the testing of reading. *Reading in a Foreign Language*, 7.1: 511-517.

Meyer, B.J.F. 1975. *The organization of prose and its effects on memory*. Amsterdam: North-Holland Publishing Co.

Meyer, B.J.F. 1985. Prose analysis: purpose, procedures, and problems: Parts I & II. In Britton, B.K. & Black, J.B. (eds.) *Understanding Expository Text*. Hillsdale, N.J.: Lawrence Erlbaum, 11-64, 269-304.

Meyer, B.J.F. 1987. Following the author's top-level organization: An important skill for reading comprehension. In Tierney, R.J., Anders, P.L. & Mitchell, J.N. (eds.)

Understanding readers' understanding. Hillsdale, N.J.: Lawrence Erlbaum, 59-76.

Meyer, B.J.F., Brandt, D.M. & Bluth, G.J. 1980. Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, 16.1: 72-103.

Meyer, B.J.F. & Freedle, R.O. 1984. Effects of discourse type on recall. *American Educational Research Journal*, 21.1: 121-143.

Meyer, B.J.F., Haring, M.J., Brandt, D.M. & Walker, C.H. 1980. Comprehension of stories and expository text. *Poetics*, 9.2: 203-211.

Meyer, B.J.F., Marsiske, M. & Willis, S.L. 1993. Text processing variables predict the readability of everyday documents read by older adults. *Reading Research Quarterly*, 28.3: 234-249.

Meyer, B.J.F. & McConkie, G.W. 1973. What is recalled after hearing a passage? *Journal of Education Psychology*, 65.1: 109-117.

Meyer, B.J.F. & Rice, G.E. 1984. The structure of text. In Pearson, P.D. (ed.) *Handbook of reading research.* New York: Longman, 319-352.

Munby, J.L.1978. *Communicative Syllabus Design.* Cambridge: Cambridge University Press.

Pearson, P.D. & Johnson, D.D. 1978. *Teaching reading comprehension.* New York: Holt, Rinehart and Winston.

Perkins, K. & Brutten, S.R. 1988. An item discriminability study of textually explicit, textually implicit, and scriptally implicit questions. *RELC Journal*, 19.2: 1-11.

Rosenshine, B.V. 1980. Skill hierarchies in reading comprehension. In Spiro, R.J., Bruce, B.C. & Brewer, W.F. (eds.) *Theoretical Issues in Reading Comprehension.* Hillsdale, N.J.: Lawrence Erlbaum, 535-554.

Rost, D.H. 1993. Assessing the different components of reading comprehension: fact or fiction. *Language Testing*, 10.1: 79-92.

Skehan, P. 1988/89. Language testing: survey article Parts 1 & 2. *Language Teaching Abstracts*, 21.4:211-221 & 22.1:1-13.

Shohamy, E. & Inbar, O. 1991. Validation of listening comprehension tests: the effect of text and question type. *Language Testing*, 8.1: 23-40.

Urquhart, A.H. 1987. Comprehensions and interpretations. *Reading in a Foreign Language*, 3: 387-409.

Urquhart, S. & Weir, C.J. 1998. *Reading in a Second Language: Process, Product and Practice.* Harlow: Longman.

Vipond, D. 1980. Micro- and macroprocesses in text comprehension. *Journal of Verbal Learning and Verbal Behavior*, 19.3: 276-296.

Weir, C.J., Hughes, A. & Porter, D. 1990. Reading skills: hierarchies, implicational Relationships and identifiability. *Reading in a Foreign Language*, 7.1: 505-510.

Appendix

International Aid

The industrialized countries between them possess 78% of all existing wealth in the world. This means that the other countries, which are usually called the 'Third World', have about 22%, even though their population is about 76% of the world's total. Many rich industrialized countries give aid to poorer Third World countries. The intention is simple – giving aid in this way should help the poorer countries to improve their situation. Of course they hope that aid will no longer be necessary in the end, since the Third World countries will have become able to look after themselves.

However, many people argue that much of the aid given to Third World countries does more harm than good. One example of this is 'tied aid'. Money or machinery is given to a Third World country, but on certain conditions. These usually mean, for example, that the receiving country has to spend the money on what is produced in the giving country. As a result, the Third World country may have to buy products it does not need, or at a higher price.

At the same time Third World countries become dependent on industrialized countries. They need them more and more. For

example, a Third World country may be given expensive tractors. Agricultural productivity may improve enormously, but when the tractors go wrong, they will require skilled mechanics or expensive spare parts. Either way, the poor country needs to pay money to the richer country to repair the tractors.

Moreover, most aid has been used in cities. This makes life there look more attractive, offering jobs which are highly paid and which are not available in rural areas. So people leave the countryside and move to cities. As a result, the countryside becomes empty and the country can no longer produce enough food for its people. At the same time, cities become overcrowded and there are all sorts of problems, from housing shortages to poor health facilities. Worse still, there may not be enough jobs for all the people who come to the cities hoping that they will become richer; many of them, in fact become poorer than before.

Graduate School of Language Sciences
Kanda University of International Studies
1-4-1 Wakaba, Mihama-ku, Chiba-shi,
Chiba-ken 261-0014
JAPAN

m.kobay@kanda.kuis.ac.jp