Development of Multiple Choice Items for Reading Comprehension Assessment

Lance Burrows

Abstract Creating reading comprehension assessment instruments remains a problem for many EFL educators and researchers, alike. This paper will first provide an overview of literature pertaining to the various facets that make up the concept of reading comprehension. In addition, the cognitive processing model framework proposed by Embretson and Wetzel, will be discussed in detail. From the basis of this model, an explanation of how to create reading comprehension multiple-choice items of varying difficulty will be discussed.

Keywords: reading comprehension, language assessment, test development, cognitive psychology, educational psychology

読解力テストにおける多肢選択項目開発

ランス　ブロース

要旨　読解力を測る問題を作成することは、多くの外国語教師および研究者に依然として大きな課題を残している。この研究では、最初に、読解力という概念を構成する多様な側面について書かれた先行研究の概要を述べる。さらに、EmbretsonとWetselによって提唱された認知処理モデルの枠組みを詳細に論じる。このモデルを土台として、様々な困難を伴う読解力テストの多肢選択項目作成方法を論じる。

キーワード：読解、言語能力評価、テスト開発、認知心理学、教育心理学

Introduction

For students of English as a Foreign Language (EFL) in Japan, it can be a huge struggle to attain quality input to further their language development. Without sufficient exposure to English, the learning process can be a series of daunting tasks. Consequently, many students turn to reading English as one way of gaining access to the language. Through books, magazines, newspapers, Internet material, and other sources of written input, students are given an open door to the language.

One of the main goals for any EFL reading teacher is to provide pupils with the opportunities and challenges to best hone their reading skills. One of the problems that many reading teachers face, however, is how best to assess reading comprehension. This assessment assists teachers in learning what further skills students need to acquire in order to become more proficient readers. In addition, proper assessment can help learners gain insight into their own strengths and weaknesses in order to improve their study. As a result, many teachers attempt to develop their own set of reading comprehension tests that often contain multiple-choice questions. Many of these tests, however, fall short of actually providing learners with a useful, accurate gauge of their reading ability due to an inherent problem with the items themselves.

Creating an assessment measure that will include items of varying levels of difficulty is a particularly challenging task for many educators. This paper will review relevant literature from the field of reading comprehension and provide a comprehensive overview of the pertinent components of multiple-choice test development. In particular, this paper will instruct readers on how to vary the level of difficulty of multiple-choice items on a reading comprehension test.

Reading Comprehension Based on Anderson's Cognitive Theory of Learning

In order to begin to understand how to develop test items that will better assess reader comprehension, a brief overview of reading as a cognitive process must first be provided. Two fundamental principles lying at the core of this cognitive view of human thought and action are: (a) human behavior is directed by individuals' perceptions and interpretations of their own experiences, and (b) the manner in which individuals think and reason often resembles the way in which computers process information (Shuell, 1986). Individuals are said to "process" information through the interplay between the environment and the mind. The thoughts involved in this cognitive activity are referred to as "mental processes."

Reading comprehension is viewed as comprising active and complex processes in which individuals formulate meaning from written information (Anderson, 1985; Howard, 1985; Pearson, 1985). Anderson has separated the comprehension process into three distinct stages: *perceptual processing, parsing, and utilization*. Although the word "stage" may imply a linear relationship between these comprehension processes, they, actually, may be described as recursive; movements from one process to the next and then back to the previous one may occur.

In reading comprehension, *perceptual processing* entails attention being focused on written text, with parts of the written input being retained in short-term memory. Due to the limitations on short-term memory, incoming information will undergo some preliminary analyses after which the majority of the information will be expelled and replaced by incoming new information. However, encoding processes may convert some of the text to meaningful representations even at this stage.

In *parsing*, words and phrases from the text are used to make meaningful mental representations of the material being read. This includes *decoding*, which means readers match the visual pattern of a word with a representation in declarative knowledge stored in long-term memory. It does not include comparing the meaning of the new mental representation with that which has been accessed in long-term memory. That will come later.

The third process, *utilization*, also referred to by some researchers as *elaboration* (Gagne, 1985), entails relating a mental representation of the text meaning to declarative knowledge in long-term memory. This process takes the mental representation that has been matched with those in declarative knowledge from the *parsing* process and discovers the meaning related to that representation in long-term memory. This interplay between information that we already know and information that is completely new is the basic determinant for reading comprehension.

Taxonomies of Test Development

On reading comprehension tests, there are certain factors that affect the difficulty of multiple-choice items. These factors can be exploited to create a set of questions of varying difficulty levels which will thereby more accurately measure the abilities or knowledge of a testing population.

Firstly, the types of questions used on the test play a role in modifying item difficulty. Early taxonomies have attempted to classify question types (Barrett, 1979;

Bloom, 1956). In these early taxonomies, question types were basically differentiated based on the level of mental energy expended to reach an answer. Those items requiring simple recognition of passage information would be seen as less demanding than those items which involved higher levels of inferencing and problem solving. Although some scholars would debate this, it is usually found that items demanding these higher levels of mental processing are more difficult to answer.

Pearson and Johnson (1978) created another taxonomy that included three categories; text-explicit, text-implicit, and script-based questions. Text-explicit questions were those where both question information and the correct answer are found in the same sentence. Text-implicit items refers to items where the answer and the question information are in two different sentences, requiring the test-taker to integrate at least two pieces of information. The last category, script-based questions are those that require readers to integrate text information with their background knowledge because answers cannot be found in the text itself. Test items are considered to increase in difficulty as the content needed to answer each particular item moves away from the content of the text.

Embretson and Wetzel's Cognitive Processing Model Framework

As can be seen by the above explanation, reading comprehension is an intrinsically complex, multi-dimensional skill that is relatively difficult to assess. When creating instruments meant to measure reading comprehension, there are certain considerations that must be taken. There are specific assessment challenges caused by the interaction between text, task, and reader that need to be addressed. In line with the cognitive features of Anderson's cognitive-based theories of learning, there is a cognitive processing model framework developed by Embretson and Wetzel (1987) from which multiple-choice item features (e.g., type of questions, passage length, etc.) can be analyzed and their contribution to the difficulty of items can be assessed.

Embretson and Wetzel's model (1987) of cognitive processing of reading comprehension has been a tool used by many researchers to investigate various aspects of reading comprehension tests and test items. The model accounts for basically two general processes: text representation and response decision. As its main premise, those items or texts which require more cognitive processing will be more difficult.

Embretson and Wetzel claim there are three decision processes involved in reading comprehension multiple-choice tests: encoding and coherence, text mapping, and

evaluating truth status of the response alternatives. Encoding, similar to Anderson's perceptual processing, is the first stage of cognitive processing, where attention is focused on written text and portions of the text are retained in working memory. Coherence, much like Anderson's parsing, involves readers starting to create mental images of the information being read. Also in this stage, parts of these images are teamed up with information in long-term memory.

In regards to item difficulty, encoding and coherence processes become more difficult when vocabulary in the questions or the response alternatives is challenging to comprehend. In other words, high levels of vocabulary terms strain readers' working memory, stealing cognitive energy that could be used to answer the question or revisit parts of the text for further investigation. The difficulty of coherence processes for the items and the response alternatives lies in the influence exerted by propositional density of the words in the items or response alternatives. Propositional density refers to the ratio of the number of propositions to the length of the item or response alternative. When propositional density is high, these items are difficult to process for later recall and comprehension, caused by the limitations of working memory capacity. Embretson and Wetzel "found that the propositional density of the passage and the extensiveness of the reasoning required to map the question and answer onto the passage were the two main factors that influenced item difficulty" (Ozuru, et al., 2008, p. 1003). Propositional density has been shown to influence the processing demands of a text (Kintsch & van Dijk, 1978) and item difficulty (Embretson & Wetzel, 1987). As such, longer sentences require more mental effort (working memory) to process, so sentences with more and longer words are regarded as more difficult.

As a second component of decision processes in Embretson and Wetzel's model, the text-mapping stage refers to readers' relating the propositions in the question and response alternatives to the information retrieved from the passage. Difficulty in text-mapping is affected by the amount of information required from the passage to answer the question. This hypothesis falls in line with the taxonomies mentioned above (Barrett, 1979; Bloom, 1956; Pearson & Johnson, 1978). Simple recognition of passage information would be much easier than inferencing information that was not directly attainable from the text (Barrett, 1979; Bloom, 1956). When question information and correct answers are found in the same sentence, the information needed from the passage to answer the item is relatively minimal, therefore theoretically easier to answer than a question which requires one to look at more than one sentence or into one's

own background knowledge to discover an answer (Pearson & Johnson, 1978). An underlying assumption of all of these taxonomies in regards to item difficulty is that the more cognitive processing involved in gleaning an answer from the text, the more difficult an item would be. Embretson and Wetzel (1987) postulate that, as the amount of text required for answering a question increases, so does item difficulty.

The third and final component of the model, evaluating truth status, comprises a two-stage process of falsification and confirmation of response alternatives. In Embretson and Wetzel's study (1987), this two-stage process was considered the strongest predictor of item difficulty. These processes detail the degree to which material from the passage could be used to make decisions about the available response alternatives. Items where the correct response could be directly confirmed by the text required minimum processing. Those items that included distracters that could be clearly contradicted directly from the material in the passage also required limited processing.

In addition to the matching between text and the alternative responses in a question, the level of difficulty of the vocabulary used in the response options also was found to affect item difficulty. In their study, Embretson and Wetzel (1987) found that distracters that contained difficult vocabulary were less likely to be processed as a potential correct response, thereby reducing the processing load and making the item easier, overall. On the other hand, it was also found that when the correct answer contained difficult vocabulary, it was much more difficult to confirm and therefore required higher mental processing, making it a more difficult item.

Factors in Addition to Embretson and Wetzel's Cognitive Processing Model Framework

Along with the vocabulary level of the questions and the various response alternatives, the phrasing and ordering of the information in the questions and response alternatives has also been shown to affect item difficulty. A question which is phrased in the same way as the corresponding answer in the text will require less cognitive energy than a question in which neither the order nor the wording of the information in the question matches that of the text in the passage. Items that entail transformed paraphrase questions are considered difficult because ideas in the text must be reworded and reordered to match the question to the location of the information needed to correctly answer it (Craik & Lockhart, 1972).

Furthermore, Sheehan and Ginther (2001) also hypothesized that information

found earlier in a passage was more readily accessed than that found later in a text. In other words, creating items which require readers to access information later in the passage in order to falsify incorrect answers or confirm correct ones would be more cognitively taxing for the test-taker, and therefore more difficult.

Finally, the differences between concreteness and abstractness of required information to answer questions will also affect the difficulty of items. In Mosenthal's (1996, p. 323) scheme, he identified five levels of abstractness/ concreteness. The levels are 1) most concrete, the "identification of persons, animals, or things," 2) highly concrete, the "identification of amounts, times, or attributes," 3) intermediate, the "identification of manner, goal, purpose, alternative, attempt, or condition," 4) highly abstract, the "identification of cause, effect, reason, or result," and 5) most abstract, the "identification of equivalence, difference, or theme." In this scheme, Mosenthal postulated that "the more concrete the requested information is, the easier it is to complete document tasks (1996, p. 327)."

Summary

In summary, the following factors will, theoretically, lend to item difficulty.

1. Using more difficult vocabulary in the question or in the response alternatives

2. Creating questions and response alternatives that contain higher propositional density

3. Asking questions of which the answers are not directly written in the text, requiring test-takers to infer meaning from contextual clues in the text

4. Asking questions of which the answers are found in more than one part of the text, requiring test-takers to synthesize two or more pieces of information to answer the question (the further away from each other in the passage these pieces of information lie, the more difficult the question will be)

5. Asking questions of which the relevant information that may be used to falsify distracters or confirm the correct answer is limited

6. Using more difficult vocabulary in the correct answer alternatives and using slightly easier vocabulary in the distracters

7. Re-ordering or re-phrasing the words in the response alternatives so that they are different from the corresponding material directly found in the text

8. Creating questions of which the relevant information that may be used to falsify distracters or confirm the correct answer is located near the end of the text

9. Developing questions of which the relevant information that may be used to falsify distracters or confirm the correct answer is relatively abstract, instead of concrete

Anderson posits that reading comprehension is an intricate series of processing tasks that require many levels of attention on the part of the reader. By taking the above nine factors into account when creating reading comprehension assessment instruments, educators can better capitalize on these facets of difficulty in creating a more well-rounded, balanced, and all-encompassing instrument that will measure more accurately, the varying levels of the knowledge base of the learners being assessed.

References

Anderson, J. R. (1985). *Cognitive psychology and its implications.* New York: Freeman.

Barrett, T. (1979). Taxonomy of reading comprehension. In R. Smith & T. Barrett (Eds.), *Teaching reading in the middle grades.* Reading, MA: Addison-Wesley.

Bloom, B. (1956). *Taxonomy of educational objectives: Handbook 1, Cognitive domain.* New York: MacKay Publishers.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent models for paragraph comprehension. *Applied Psychological measurement, 11*, 175-193.

Gagne, E. D. (1985). *The cognitive psychology of school learning.* Boston, Mass.: Little, Brown.

Howard, D. (1985). *Cognitive psychology.* New York: Macmillan.

Kintsch, W., & van Dijk, A. (1978). Toward a model of text comprehension and production. *Psychology Review, 85*, 363-394.

Mosenthal, P. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology, 88*, 314-332.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40*(4), 1001-1015.

Pearson, P. D. (1985). Changing the face of reading comprehension instruction. *The Reading Teacher 38,* 724-38.

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

Sheehan, K., & Ginther, A. (2001). *What do multiple choice verbal reasoning items really measure? An analysis of the cognitive skills underlying performance on a standardized test of reading comprehension skill.* Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56*(4), 411-436.