Current Issues in Education
Mary Lou Fulton Teachers College • Arizona State University
PO Box 37100, Phoenix, AZ 85069, USA

Volume 17, Number 1 February 04, 2014 ISSN 1099-839X

Assessment of a Reading Comprehension Instrument as It Relates to Cognitive Abilities as Defined by Bloom's Revised Taxonomy

Lorraine Dagostino, James Carifio, Jennifer D. C. Bauer, and Qing Zhao
University of Massachusetts at Lowell

Nor H. Hashim
Universiti Sains Malaysia

More often than not, the assessment of literacy has focused upon how well readers attain various levels of reading comprehension or demonstrate proficiency with specific reading skills rather than reveal a reader's cognitive abilities as reflected in conceptual frameworks such as Bloom's Revised Taxonomy (Anderson, et al., 2001). Where tests of reading comprehension have classified test items by type for the purposes of item analysis, it usually is only by reading skill. The focus of this study is to change from this functional approach to one that examines how well, if at all, a test created in Malay, and translated into English, that was developed for primary and intermediate grade readers, may also be able to determine cognitive levels of understanding as described by Bloom's Revised Taxonomy – The Cognitive Dimension (Hashim et al., 2006). To accomplish this task, the raters evaluated the questions from both Malay tests using Bloom's Revised Taxonomy of Cognitive Abilities. By using the Bloom's Revised Taxonomy as a system of classification, the researchers were able to more accurately pinpoint the specific cognitive abilities being assessed by each test item. These findings suggested that classification by cognitive level allows one to measure specific cognitive abilities as defined by Bloom's Revised Taxonomy. This is significant because Bloom's Revised Taxonomy gives us objectives for classifying the learning, teaching and assessing of the cognitive dimension of thought that is central to instruction in most subject areas, and in relationship to our work in reading comprehension as an aspect of assessment of literacy in a way that differs from most current measures of reading comprehension.

Keywords: reading comprehension; cultural background; cognitive process; Bloom's Revised Taxomony

More often than not, the assessment of literacy has focused upon determining how well readers attain specific levels of reading comprehension, such as literal meaning, inferential meaning or applications of what is understood, or how well readers demonstrate the attainment of particular reading skills, such as decoding, reading for the main idea and significant details, the tone of a passage, or drawing conclusions (Cain & Oakhill, 2006; Storch & Whitehurst, 2002). Where tests of reading comprehension have classified items by type for the purposes of scoring or/and item analysis it usually is only by reading skill type, and thus operationally such tests are *skill views* and *skill definitions* of reading comprehension. (By *skill views* and *skill definitions* we refer to the theories that define reading comprehension as a set of particular skills – such as decoding, letter knowledge, and

phoneme awareness – that must be mastered and applied in order for the reader to comprehend text.) This functional approach has been helpful, but represents only one view of reading comprehension and the reader's ability to read either fiction or non-fiction with a range and various degrees of understandings. A different view and perhaps more significant concern is how the reader's general (qualitative) cognitive processing abilities, as characterized by say Bloom's Revised Taxonomy of the Cognitive Dimension (Anderson et al., 2001), relate to reading comprehension, and how well, if at all, these general cognitive processing abilities (hereafter termed "cognitive abilities" in this article) are actually being considered in the assessment instrument when determining how well the reader ascertains the meaning of text (McNamara & Kendeou, 2011; Winstead, 2004). Although there is a sense in the research literature that a reader's cognitive abilities influence reading comprehension levels and abilities, little work has been done in the last few decades to determine how well assessment instruments for measuring reading comprehension evaluate the cognitive abilities that are being assessed by individual test items, as it is not a view that is consciously attended to and allowed to emerge explicitly in the skill views and definitions of reading comprehension. Obviously, both views are needed, and explicitly knowing both views and their relationships for a given reading comprehension instrument would give a far more powerful and comprehensive characterization of that instrument, as well as reading comprehension *per se* and research done on reading comprehension in general and specifically. Also, being able to characterize a given set of reading comprehension test items according to multiple views and theoretical frameworks would also allow far more efficient and powerful research designs and study, and thus our interest in this topic and problem.

Previous Work

Previously, we completed a study evaluating the comparability of two reading comprehension tests written to assess the reading ability of Malaysian primary (1-3) and intermediate (4-6) grade readers with its English translation in terms of reading skills and reading comprehension levels according to a conceptual framework of reading comprehension developed by Dagostino and Carifio (1994) that proved to be quite useful for evaluating the nature of the test items in reading comprehension instruments (see Dagostino, Carifio, Bauer, & Zhao, 2013). That research was able to establish strong correlations across the two versions of the tests on the classification of reading skills and levels of reading comprehension.

The original tests were written in Malay for a nationwide study of reading comprehension in Malaysia. We then translated both tests into English for the purpose of determining the relationship of classifications of reading skills and levels of reading comprehension in two different languages. While the tests were not constructed with the cognitive dimensions of thought as part of the test item specifications (i.e., Bloom's general cognitive [processing] abilities), we thought that this factor would be an interesting dimension to explore in further work, and thus became the focus of the present study.

Present Work

The present study extended our previous study to examine the English version of the above described test further to see if this version of the test in any way could be reasonably and usefully characterized in terms of general cognitive processes and abilities. Therefore, the purpose of the present work was to determine if test items on the English versions of Malay tests reflected the categories of general cognitive abilities described in Bloom's Revised Taxonomy of the Cognitive Dimension (Anderson et al., 2001) by evaluating inter-rater judgments of the 100 test items on two tests developed for evaluating the reading comprehension abilities of primary (grades 1-3) and intermediate (grades 4-6) grade readers in Malaysia. Consequently, with this purpose in mind, we set out to explore the following research question:

> What is the inter-rater agreement for each test item classified using Bloom's Revised Taxonomy of the Cognitive Dimension (Anderson et al., 2001) when the judgments of all three raters are analyzed as a group?

Organization of the Article

With the above goals mind, we will begin with a description of the Malay tests, their development, and the work of the present study. Then, Bloom's Revised Taxonomy of the Cognitive Dimension (Anderson et al., 2001) will be described, as well as its application to the present study. Next, we detail the components of the study itself, including the parameters and limitations, methodology, procedures, results and subsequent data analysis. We then finish with an overall summary and final comments on the implications of this work.

The Malay Tests

This next section of the paper, describing the Malay Tests, was originally published as part of the author's previous study (see Dagostino, Carifio, Bauer, & Zhao, 2013).

The Description and Construction of the Malay Tests

The original two Malay tests, constructed by a team of researchers at the Universiti of Sains Malaysia, were developed for the purpose of evaluating reading comprehension abilities of students in the primary grades (Test I for grade 1-3, Test II for grades 4-6) in Malaysia (NorHashim, 2004). The following section of this article describes the process for the development and the content of these tests.

Steps for Design and Content of the Malay Instruments

Using the Dagostino-Carifio model (1994) of reading comprehension as a theoretical basis, the

Table 1

Table of Specifications for Malay Reading Comprehension Tests with this general Template being the same for Test I and Test II

Reading Comprehension Category	Code	Reading Skills
Literal (L)	L1A, L1B, L1C	identifying meaning of word/ phrase/ sentence
	L2	identifying main idea
	L3	identifying important point
	L4	making comparison
	L5	identifying cause-effect
	L6	identifying sequence of ideas/events
Inferential (F)	F1	interpreting main idea
	F2	interpreting important point
	F3	interpreting comparison
	F4	interpreting cause-effect
	F5	making a conclusion
Critical Creative (K)	K1	evaluating
	K2	making a conclusion
	K3	internalizing
	K4	identifying the moral of the story/lesson

development of the test focused on three components: a) defining and selecting the category of the comprehension level as well as of the comprehension skill, b) selection and development of the reading texts, and c) the development of the test questions and the answers. The two tests were designed by conducting a preliminary survey that included a discussion with Malay teachers, a review of teaching learning materials and observations of teachers teaching in a classroom. Once the survey was completed, a first draft was developed for Test I and for Test II. The writing of the first draft was accomplished through a series of workshops with Malay language teachers, experts from Curriculum Development Center, administrators from the District Education Office and State Education Department, lecturers of School of Educational Studies from the Universiti Sains Malaysia (NorHashim, 2004). As a result of this work, the researchers established the following Table of Specifications (Table 1), which outlines the relationships between the reading comprehension levels and reading skills underlying the construct of both tests.

Defining the Reading Comprehension Levels and the Reading Comprehension Skills

The reading comprehension levels and the reading skills determine the difficulty and the nature of the reading texts and the test items. The Malaysian tests have three comprehension levels defined as follows (NorHashim, 2004):

(a) *Literal (message extraction) Reading Comprehension, which* refers to the memorization of facts in texts where information is explicitly stated at a basic level of thinking;

(b) *Inferential (message interpretation) Reading Comprehension* which refers to the ability of students to interpret meaning where they need to use overt information along with intuition, reasoning, and experience to attain the higher level of thinking assessed by the Malay tests; and,

(c) *Critical/Creative (message evaluation) Reading Comprehension*, which refers to the student's ability to do an overall critical evaluation of certain information or an idea that has been read in terms of the precision and/or suitability of the given information of a new idea, encountered. This critical evaluation may require some divergent thinking and depend to some degree upon the knowledge and personal experience of the student, but it focuses mostly on convergent critical thinking being done by the student.

Reading comprehension skills. There are ten reading comprehension skills that are assessed by the Malay tests (NorHashim, 2004): (a) identifying meaning of word/phrase/sentence; (b) identifying the main idea; (c) identifying the important point; (d) identifying the cause-effect relationship; (e) identifying the sequence of ideas/events; (f) making a comparison; (g) drawing a conclusion; (h) evaluating; (i) internalizing; (k) identifying the moral of the story/lesson. These ten skills range from simple reading comprehension to what is called deep or deeper understanding, which is a first step towards what is called evaluative reading. These skills are the ones that usually constitute the classification of items assessed in most reading tests.

Types and Contents of Reading Texts

There are several types of text that make the text broad in scope and representative of various types of reading of non-technical materials that are encountered in daily reading situations (NorHashim, 2004). There are essays, fiction, reports, letters, poems, biographies, speeches, dialogues, and news reports. There are 12 texts for Test I and 12 Texts for Test II. There are various subjects (literature, history, etc.) The individual texts are 100 words of less for Test I and 100 words or more for Test II. The passages in the test for grades (1-3) are simpler in structure as well as expectations for level of reading comprehension than those used for grades 4-6. A research group, 3 expert teachers, teacher trainers, psychometric and experts from the university developed the texts, with ideas for the texts coming from books and magazines.

Development of the Test Items

The question and answer formats for the tests took various forms such as a) sentences from text that needed completion with a choice of answers, b) items that needed a choice of answers in multiple choice form, and c) instructions and blanks to be filled in with multiple choice form. An item specification table was developed to categorize the types of items in the test (NorHashim, 2004). Each test consists of 50 multiple choice items designed to evaluate reading comprehension with consideration given to reading skill ability and reading comprehension level. Some specific things were considered in the item development. They are as follows: a) arrangement of each item was based upon reading comprehension skill (forms, style, pupils' existing knowledge), and b) implicit information and inferential definition. In the case of implicit information, the text considers information in the text and students' background. In the case of inferential definition the test considers an integrated synthesis of literal with existing knowledge, intuition and reader's imagination.

The following two Table of Specifications include the classification by reading comprehension level and reading skill for each test item. Both Malay tests were built from the same general Table of Specifications, but classification by reading skill varied for each test (Table 2 and 3).

Table 2

Malay Table of Specifications Including Test Items by Classification for Test I

Reading Comprehension Level	Code	Reading Skills	Item Numbers
Literal (L)	L1A, L1B, L1C	identifying meaning of word/ phrase/ sentence	8
			13
			41
			46
	L2	identifying main idea	1
			5
			9
			47
	L3	identifying important point	2
			6
	L4	making comparison	10
			14
			15
	L5	identifying cause-effect	3
			7
			11
			42
	L6	identifying sequence of ideas/events	4
			12
			16
Inferential (F)	F1	interpreting main idea	21
			25
			29
			43

	F2	interpreting important point	17
			26
			48
	F3	interpreting comparison	18
			22
			30
			31
			49
	F4	interpreting cause-effect	19
			23
			27
			32
	F5	making a conclusion	20
			24
			28
			44
Critical Creative (K)	K1	evaluating	33
			40
			45
	K2	making a conclusion	34
			37
			38
	K3	internalizing	35
			39
	K4	identifying the moral of the story/lesson	36
			50

Table 3

Malay Table of Specifications Including Test Items by Classification for Test II

Reading Comprehension Level	Code	Reading Skills	Item Numbers
Literal (L)	L1A, L1B, L1C	identifying meaning of word/ phrase/ sentence	1
			5
			41
			46
	L2	identifying main idea	9
			13
			42
	L3	identifying important point	2
			10
			47
	L4	making comparison	6
			14
			15
	L5	identifying cause-effect	3
			7
			11
			16
	L6	identifying sequence of ideas/events	4

			8
			12
Inferential (F)	F1	interpreting main idea	17
			21
			25
			29
	F2	interpreting important point	18
			22
			26
			30
	F3	interpreting comparison	27
			31
			43
			48
	F4	interpreting cause-effect	19
			23
			32
			44
	F5	making a conclusion	20
			24
			28
			49
Critical Creative (K)	K1	evaluating	33
			38
			50
	K2	making a conclusion	34
			37
			39
			40
	K3	internalizing	35
	K4	identifying the moral of the story/lesson	36
			45

Design and Choice of Answers and Distracters

A multiple-choice format was used because it was considered as most objective. Each answer had 4 options (A, B, C, D for each item with each option coded A=1, B=2, C=3 and D= 4). The correct answer was scored 1, and the wrong answer was scored 0. The design of the answers and distracters required a) the suitability of choice of answers relative to the cognitive task that was related to the content and the texts, and b) syntax and semantic forms needed to be different from the texts so that students could be assessed on how well they understood the context of the meaning (NorHashim, 2004).

Reliability Measures of the Two Malay Tests

The Malay researchers examined three types of internal consistency reliability estimators for both tests with the results being almost identical for both tests. The first internal consistency (of test-taker overall performance) reliability estimator computed was the Cronbach alpha coefficient, which was r=+.66 (N=2763) for Test I and r=+.61 (N=4101) for Test II. As is well known, test length, sample size, and test content and item type *heterogeneity* affect and limit the size of the Cronbach alpha one will observe in any given context. As test content and the cognitive levels and operations assessed are so heterogeneous for both tests, the Cronbach alphas observed for each test are quite good to excellent given that test lengths (50 items each) and sample sizes (N=2763 and 4101+) and are in the range that one would expect given the qualitative characteristics of both tests.

The second internal consistency reliability estimator the Malay researchers computed was the Guttman reliability coefficient, which assess the degree to which students' performances on the test are hierarchical in character (i.e., students who do well on low level items are not doing well on high level items and vice versa), which performances should be for Test I and Test II given how they were constructed and their qualitative characteristics. The Guttman reliability coefficient for Test I was r=+.77 (N=2763) and for Test II was r=+.72

(N=4101), which are excellent to outstanding and indicate that this particular qualitative characteristic of both tests are as hypothesized and purported.

The third internal consistency reliability estimator the Malay researchers computed was the Kuder-Richardson odd-even items reliability coefficient, which assess the degree to which items types and their characteristics are *evenly balanced* across the test, as well as students' performances on the items on the test. For example, the Kuder-Richardson reliability coefficient would be low if all of the odd items were easy (or recall) items and all of the even items were difficult (or skill) item, or if all of the poorly constructed and non-functioning items were easy items as opposed as opposed to this characteristic being evenly balanced across both the odd and even items. The Kuder-Richardson odd-even items reliability coefficient for Test I was r=+.77 (N=2763) and for Test II was r=+.73 (N=4101), which are good to excellent and indicate that the various types of items and their various characteristics were *evenly balanced* across each test as were student performances.

As one administration internal estimates of various types of consistencies in student performances across each of these two tests and thus internal consistency reliabilities estimates, the results obtained by the Malay researchers of the three different indicators of internal reliabilities estimates were excellent. High one-time internal consistency estimates of reliabilities, however, are no guarantee that test-retest reliabilities will be equally high as they could actually be lower or higher which is why the Malay researchers are currently collecting the data to generate the test-retest reliability coefficients as these coefficients are key in the assessment of change across time on these measures. But to date, the reliabilities estimates for each test that are available are excellent and particularly so given the internal complexity of each test, and each is also initially supportive empirically of specific aspects of the construct validity of each test, although not as direct or strong evidence as other analyses might indicate.

Bloom's Revised Taxonomy: the Cognitive Dimension (Anderson et al., 2001) and its Application to the Present Study

Bloom's original taxonomy was designed to help teachers establish objectives for instruction, learning and assessment. This revised taxonomy has served to guide the design and the implementation of accountability programs and standards-based curriculum. The revision of the original taxonomy that was in the present study has been refined to incorporate new knowledge into the original framework. This revised taxonomy gave us a good conceptual framework for determining the cognitive levels and ability reflected in test items on the reading comprehension test that the researchers' expect to use as an assessment instrument in subsequent studies. The test already has been examined for general levels of reading comprehension and reading skills. What we hoped to accomplish in the present study was to see if the test items also reflected levels of the specific cognitive abilities as defined by Bloom's Revised Taxonomy (Anderson et al., 2001). This taxonomy was chosen from other ways to evaluate cognitive abilities because it is most applicable, familiar and understandable to the classroom teacher, yet detailed enough to give valuable insight into cognitive processes that are considered necessary to learning and to the assessment of success in instruction and learning (Anderson et al., 2001). Further work is planned to compare Bloom's Revised Taxonomy (Anderson et al., 2001) with other classification frameworks for measuring cognitive abilities as they may manifest themselves in tests of reading comprehension.

Using Bloom's Revised Taxonomy (Anderson et al., 2001) gave us a standard, well-recognized classification system for our immediate goals, and it also should be useful for guiding instruction and curriculum guidelines that may be generated by our present work. This consistency across these tasks should simplify the work of the classroom teacher and the researcher. In sum, Bloom's Revised Taxonomy gives us definitions for classifying the learning, teaching and assessing of the cognitive dimension of thought that is central to instruction in most subject areas, and in relationship to our work in reading comprehension as an aspect of assessment of literacy in a way that differs from most current measures of reading comprehension.

What follows here is a table of Bloom's Revised Taxonomy (Anderson et al., 2001), and descriptions of the categories that were used in the present study.

Table 4

Definitions of the Categories of Bloom's Revised Taxonomy – the Cognitive Dimension (Remembering, Understand, Apply, Analyze, Evaluate and Create)

Remembering	*Recognizing* involves retrieving relevant information from long-term memory in order to compare it with presented information. Also identifying
	Recalling involves retrieving relevant information from long-term memory when a prompt is given. The prompt often is a question. Also retrieving.
Understand	*Interpreting* occurs when a student is able to convert information from one representation to

	another representation. Also translating or paraphrasing.
	Exemplifying occurs when a student gives a specific example or instance of a general concept or principle. Also illustrate.
	Classifying occurs when a student recognizes that something belongs to a certain category. It is a complementary process to exemplifying.
	Summarizing occurs when a student suggest a single statement that represents presented information or abstracts a general theme. Also generalize or abstract.
	Inferring involves finding a pattern within a series of examples or instances. The student abstracts a concept or a principle that accounts for a set of instances. Also extrapolating or concluding.
	Comparing involves detecting similarities and differences between two or more objects, events, ideas or situations. Also contrasting, matching.
	Explaining occurs when a student is able to construct and use a cause-effect model of a system. The model may be derived from a formal theory or may be grounded in research and experience. Also constructing a model.
Apply	*Executing* occurs when a student routinely carries out a procedure when confronted with a familiar task. Also carrying out.
	Implementing occurs when a student selects and uses a procedure to perform an unfamiliar task. It is carried out in conjunction with understand. Also using.
Analyze	*Differentiating* occurs when there is a determination of the relevant or important pieces of a message in relation to the whole structure.
	Organizing occurs relative to the way the pieces of a message are organized into a coherent structure.
	Attributing occurs when the underlying purpose or point of view of the message is related to the entire communication.
Evaluate	*Checking* involves testing for internal consistencies or fallacies in an operation, product, or communication to see whether data support or disconfirm hypothesis or conclusions as well as the accuracy of facts.
	Critiquing involves judging a product, operation or communication against externally imposed criteria and standard.
Create	*Generating* occurs when a problem is represented and alternatives and hypothesis that meet certain criteria are produced.
	Planning occurs when a solution method is devised that meets a problem's criteria for developing a plan for solving the problem.
	Producing occurs when a plan is carried out for solving a given problem that meets certain specifications.

What creates difficulty in assessing reading comprehension is the various ways the construct of reading comprehension has been conceptualized and discussed in the research literature as well as in the way those constructs have been applied to instruction. These variations also influenced our thinking by creating an incongruence and dissonance in the results in previous work thus leading to the present study. Early in the paper we indicated that the more traditional view of reading comprehension based in a behaviorist view has driven the field of assessment of reading comprehension for some time (Cain & Oakhill, 2006; Storch & Whitehurst, 2002). Because of this influence of a behaviorist view we think that having an assessment instrument that focuses on assessing the reader's ability to get the meaning of a text has been lost. This traditional view of reading comprehension is very much reflected in part of the way the Malay researchers implemented a construct of reading comprehension through a skills perspective that is aligned with this behaviorist thinking. What this view may represent, along with the more current discussion of strategies, is how students may go through of reading the text to try get to meaning, but not what the reader actually comprehends from the text. One may think of this behavior as going through the motions of trying to comprehend rather than actually using strategies to uncover the meaning in the text. The Dagostino-Carifio model (2004) attempts to extend this view by focusing on the continuously evaluative nature of the reading process as integral to assessing reading comprehension, and looking at levels of reading comprehension. It is a more holistic view of reading comprehension reflective of

cognitive ability with the requirement that an evaluative process be considered part of the construct of reading comprehension, and in fact central to it. This reshapes the construct of reading comprehension considerably to begin to include an additional cognitive dimension that begins to suggest the cognitive abilities reflected in the third way in which the construct of comprehension may be applied to reading comprehension that we have focused upon in this study – that is a cognitive framework such as Bloom's Revised Taxonomy. While we acknowledge that the Malay test considered levels of Reading comprehension, it was not really the information that is applied to subsequent instruction. Instead the emphasis still is on skills. Again, while we acknowledge that the Malay test was developed for younger readers, we believe that some of what we are saying about assessing reading comprehension still may apply, and that the construct of reading comprehension in the Dagostino-Carifio model (2004) is a more comprehensive view.

In then moving forward by using the cognitive view reflected in Bloom's Revised Taxonomy of Cognitive Abilities, we are very much shifting the focus to the thinking processes and levels of intellectual development of the reader that focuses much more on getting meaning rather than *going through the motions* that reflect the behaviorist view. It also should be noted that the cognitive ability of understanding in the Revised Bloom's Taxonomy is only a subset of the construct of reading comprehension as it is used to direct the present work.

What we may see in the progression in these constructs of reading comprehension is a movement towards an emphasis on meaning derived from an integration and synthesis of reading skills rather than on the behaviors reflected in the reading skills. With this movement towards meaning we see the reader able to make sense of a text, both explicitly and implicitly. The Dagostino-Carifio model (2004) begins to bridge this gap from focusing on behaviors to focusing on getting meaning from the text at various intellectual and cognitive levels (see Figure 1). Further, moving to the Bloom's Revised Taxonomy makes the shift to assessing cognitive abilities an even better way to determine how well the reader has gotten meaning from text.

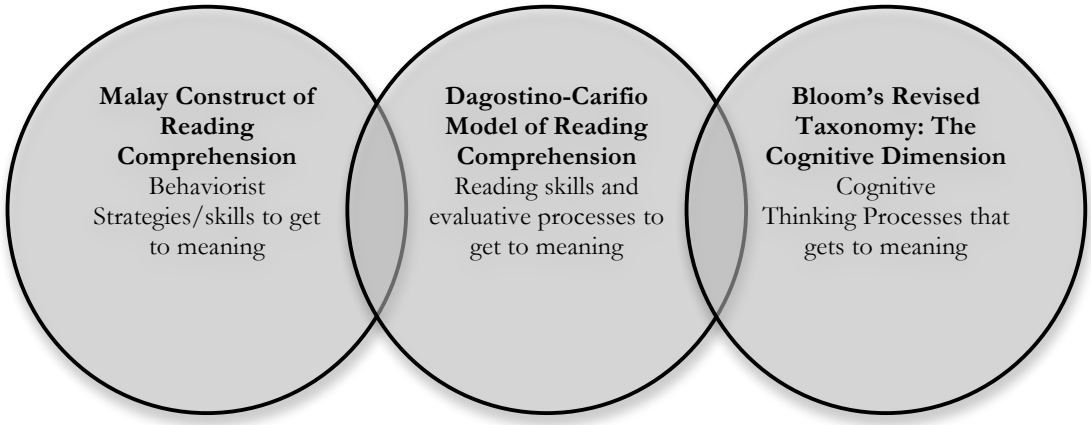

Figure 1. The Dagostino-Carifio Model of Reading Comprehension (2004) as a conceptual bridge between behaviorist and cognitive models of reading comprehension. Unlike the behaviorist approach, which emphasizes skills as a means of understanding a text, or the cognitive approach, which places emphasis on thinking processes, the Dagostino-Carifio model acknowledges a need for both reading skills and evaluative processes in or for the reader to derive meaning from a text.

This Study

Parameters and Limitations

The present work applies to the novice (grades 1-6) reader who may be approaching the early stages of formal reasoning rather than to the expert (grades 7-adult) reader who may be into the early stage of formal reasoning. The reading materials considered in this study are non-technical, such as essays, fiction, poetry, journalistic writing, rather than technical, such as expository scientific or mathematical content. The nature of cognitive thinking considered is primarily convergent, critical thinking rather than divergent, creative thinking. Lastly, the focus of this study was to determine what levels of Bloom's Revised Taxonomy of Cognitive Processes actually are reflected in these tests to see if in fact the test taps cognitive abilities in any way.

Methodology

The Translation Process

As previously stated, the original tests used in this study were developed in Malaysia by a team of researchers at the Universiti Sains Malaysia for the purpose of evaluating reading comprehension abilities of students in the primary grades (1-6) in four regions (North, East, Middle, South) in Malaysia. The two tests (Test I grades 1-3, Test II for grades 4-6) were developed for and administered to this population in a national assessment study from April to May 2004. The two tests were translated into English for our present work by a professional translation center. The original tests were forwarded intact to a native speaker of Malay who was a Communication student at the University of Massachusetts Amherst. Upon completion of the translation, a native speaker of English reviewed the text. Any revisions or questions were noted using the Track Changes feature in MS Word, and the file was returned to the original translator to either accept or reject the changes. The final file in MS Word was then submitted to the University. The lead researcher who developed the Malay version of the tests verified the accuracy and the appropriateness of the translations then reviewed the translations. The lead researcher is bilingual in Malay and English. The translations were judged by the Malaysian researcher to be satisfactory (Dagostino, Carifio, Bauer, & Zhao, 2013).

Procedures

Three of the current authors independently rated each items on the two tests according to Bloom's Revised Taxonomy as a first step in the process. The raters had either a Ph.D. in language arts and literacy, or were completing work for that degree. One of the raters spoke both English and Chinese, and another works with young children from several cultures and language backgrounds. Previous ratings by these same raters had judged the items for skills and levels as described earlier in this article with excellent results (see Dagostino et al., 2013 for details).

The three raters evaluated each test item from both Malay tests based and classified each test item using Bloom's Revised Taxonomy of Cognitive Abilities. The three expert raters completed their individual judgments by first reading each item of Test I and Test II independently, and then determining the Bloom's Revised Taxonomy level of cognitive ability they felt best applied to the dimension of reading comprehension being tested. The categories for classification were as follows: 1) Remember, 2) Understand, 3) Apply, 4) Analyze, 5) Evaluate, and 6) Create. (See Table 4 for definitions of each category).

After independent readings and ratings of the test items using the Bloom's Revised Taxonomy were completed, the raters compared their judgments with each other for all their of ratings. There was not a need for a reconciliation process among the raters based upon this discussion because of the high level of agreement among the three raters. After the quantitative analysis of the ratings was completed, the raters met again to discuss the results to evaluate the meaning of the raters' agreements on the item ratings.

Results and Data Analysis

The analysis and results section of this article presents the data and its interpretation for the research question:

> What is the inter-rater agreement for each test item classified using Bloom's Revised Taxonomy of the Cognitive Dimension (Anderson et al., 2001) when the judgments of all three raters are analyzed as a group?

The procedure used to analyze the data was the calculation of inter-rater correlation coefficients. This coefficient was computed by first getting the percentage of agreements between the three raters for a given judged (which is the explained variance) and then taking the square root of that percentage which would be the inter-rater correlation or reliability coefficient.

To judge the effectiveness of using Bloom's Revised Taxonomy as a means to classify each test item, each rater individually judged every question, and then compared their answers. The agreement rate was 98% ($r>.99$) for Test 1, and 99% for Test 2 ($r>.99$). Once these analyses were complete, the raters gathered to compile the quantitative data and examine the results. The raters' discussion on the items relative to disagreement did not show a clear pattern as to type of disagreement.

The research question addressed was, "What is the inter-rater agreement for each test item classified using Bloom's Revised Taxonomy when the judgments of all three raters are analyzed as a group?" Table 5 presents a comprehensive look at the frequencies and percentages of rater agreements about the Bloom taxonomic level of each item for both Test I and Test II. The square roots of the agreement percentages approximate the inter-rater correlation coefficients.

Table 5

Test 1: Rater Agreements

Percentages of Rater Agreements of Classification of Test Items by Bloom's Taxonomy

Type of Agreement	Frequency	Percent	Cum. Percent	r
1. Raters agreed on classification	48	.96	96	.98
2. Raters disagreed on classification	2	.04	100	
	50	100%		

Test 2: Rater Agreements

Percentages of Rater Agreements of Classification of Test Items by Bloom's Taxonomy

Type of Agreement	Frequency	Percent	Cum. Percent	r
1. Raters agreed on classification	47	.94	94	.99
2. Raters disagreed on classification	3	.06	100	
	50	100%		

As can be seen from Table 5, agreement between the raters was very high in regard to their classifications of test items by Bloom's Revised Taxonomy (r>.98). This high degree of inter-rater agreement indicates that each individual test item can be reliably classified using Bloom's Revised Taxonomy, which is a highly positive result as Bloom's revised Taxonomy can provide reading researchers with a drastically different measure of reading comprehension abilities than are traditionally assessed through skills-based only characterized reading tests. Current reading comprehension measures only include specific reading comprehension skills, whereas applying the framework of Bloom's Revised Taxonomy to the structure of a test might allow reading researchers to expand their understanding of the nature of reading comprehension, as well as to identify the potential use and application of cognitive strategies by readers during the reading process. All of these points may also apply to other tests of achievement and understanding, but further research would be needed to confirm this point empirically.

Findings and Discussion

These findings demonstrate that levels of Bloom's Revised Taxonomy of the Cognitive Dimension may reliably classify reading comprehension test items even when the items were written and validated according other views and frameworks of reading comprehension. The general Bloom characterized cognitive processes required to perform the item, therefore, are manifest in the item itself and the item's particulars, and constitute an ignored latent view (and rival theory) of the item and the test as whole similar to the unique (as opposed to the common or joint) portion of generalized tripartite item variances in factor analysis and factor analytical models of and results for instruments (see Harman, 1976 for details). This specific point means that additional unique and important information may be extracted and gained from the item and item set using this latent view or frame that will contribute significantly to finding and understandings using the test and students' performances when the test is *double scored* or *matrix scored* using the multiple frames of reference with its built in rival hypotheses tests. Such characterizations and analyses of reading comprehension tests will produce a much better, deeper and fuller understanding of reading comprehension as well as a methodology for test makers to check the quality their own work. However, there are also further benefits.

The empirical findings and facts above are also significant because Bloom's Revised Taxonomy gives us sets of explicit objectives for classifying the learning, teaching and assessing of the cognitive dimensions of thought and reading comprehension that are central and generalized to instruction in most subject areas, and in relationship to our work on reading comprehension as an aspect of assessment of literacy in a way that differs from most current measures of reading comprehension.

While many theories of reading comprehension present the process of reading as a hierarchal set of skills that are learned and applied by the reader, and thus seek to measure those particular skills (such as speed, fluency, and decoding), The Dagostino-Carifio Model (2004) diverges from this viewpoint and is most comprehensive. While acknowledging the skills view of reading and the legitimacy of identifying such skills, particularly as part of the reasoning process of reading, the Dagostino-Carifio Model (2004) proposes that reading skills are not necessarily applied in a strict sequential and hierarchical

fashion, but that they may be more fluid in nature. It also suggests that there is an evaluative process and comprehension and higher order cognitive processes that are occurring during the reading process that are more akin and better characterized by Bloom's Revised Taxonomy. Furthermore, the Dagostino-Carifio Model (2004) considers these cognitive processes and cognitive processing to be essential to understanding as well as perhaps measuring reading comprehension abilities.

As we initially stated, this study was undertaken because we believed that the traditional skill view for classifying reading comprehension test items could be significantly improved by considering a different conceptual framework for reading comprehension. The alternative view chosen focused on classifying reading comprehension test items by cognitive ability levels as well as skill levels or just either view alone. The results suggest that this alternative view is a viable and useful approach and view that may actually test reading in a more useful and comprehensive manner.

Future Research

Little has been done to examine how the measurement of the reader's cognitive abilities, as determined by Bloom's Revised Taxonomy (Anderson et al., 2001), correlates with the other characterization and classification constructs, such as those delineating reading skills or reading comprehension levels as defined by more functional views of reading for categorizing test items. In our next study, we will begin to compare and analyze the relationship, if any, between the original Malay Classification System and Bloom's Revised Taxonomy. Once that comparison is complete we hope to examine additional classification schemes for cognitive abilities to see if they too are comparable measures of performance on this instrument.

Anderson, L. E., & Karthwohl, D. (Eds.). (2001). *A Taxonomy for Learning, Teaching and Assessment.* New York: Longman.

Bloom, B. (1956). *Taxonomy of Educational Objectives: Handbook I, Cognitive Domain*. New York: Davis Press.

Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697–708.

Dagostino, L., & Carifio, J. (1994). *Evaluative Reading and Literacy: A Cognitive View*. Boston: Allyn and Bacon.

Dagostino, L., Carifo, J., Bauer, J., & Zhao, Q. (2013, February). Cross-cultural reading comprehension assessment in Malay and English as it relates to the Dagostino and Carifio model of reading comprehension. *Current Issues in Education*, 16(1), 1-16.

Harman, H. H. (1976). *Modern factor analysis.* Chicago: University of Chicago Press.

Hashim, N. H., Lah, Y. C., Ahmad, M. Z., Yaakub, R., Aziz, A. R. A., Mohamed, A. R., … Ramli, I. (2006). *Reading Comprehension in Malaysia Primary grades 1-6.* Penang, Malaysia: Universiti Sains Malaysia. (Research Grant).

McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education, 4*(1), 33 – 46.

Storch, S. A., & Whitehurst, G. J., (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38* (6), 934-947.

Winstead, L. (2004). Increasing academic motivation and cognition in reading, writing, and mathematics: Meaning-making strategies. *Educational Research Quarterly, 28* (2), 30 – 49

Article Citation
Dagostino, L., Carifio, J., Bauer, J. D., Zhao, Q., & Hashim, N. H. (2014). Assessment of a Reading Comprehension Instrument as It Relates to Cognitive Abilities as Defined by Bloom's Revised Taxonomy. *Current Issues in Education*, *17*(1). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1281

Author Notes

Lorraine Dagostino
University of Massachusetts at Lowell
O'Leary Library 519, 61 Wilder St., Lowell, MA 01854-3047
Lorraine_Dagostino@uml.edu

Lorraine Dagostino is a professor at Graduate School of Education in the University of Massachusetts Lowell. Her research interests include literacy, critical thinking and evaluative reading, theoretical issues in reading and literacy, and literature issues.

James Carifio
University of Massachusetts at Lowell
O'Leary Library 527, 61 Wilder St., Lowell, MA 01854-3047
James_Carifio@uml.edu

James Carifio is a professor in Psychology and Research Methods at the University of Massachusetts Lowell. He has done extensive research and publishing in the respective fields.

Jennifer D. C. Bauer
University of Massachusetts at Lowell
jennifer_bauer@student.uml.edu

Jennifer D. C. Bauer teaches digital media and visual art at Lowell High School, and English at the Phillips Academy, Andover Summer Session. Additionally, she is a doctoral student at the University of Massachusetts at Lowell, pursuing her Ed.D. in Language Arts & Literacy. Her research interests include Visual Arts and Literacy, Digital/New Literacies, Language Variation, Urban Education, and Culturally Responsive Pedagogy.

Qing Zhao
University of Massachusetts at Lowell
O'Leary Library 512, 61 Wilder St., Lowell, MA 01854-3047
Qing_Zhao@uml.edu

Qing Zhao recently earned her Ed.D. in Language Arts & Literacy from the Graduate School of Education, University of Massachusetts Lowell in 2013. Her research interests currently include the study of teaching and learning English as a foreign/second language (EFL/ESL) and development of EFL/ESL learners' language proficiency in both North American and Asian learning context.

Nor Hashimah Hashim
Universiti Sains Malaysia
shimah@usm.my

Nor Hashimah Hashim is a professor of education at the Universiti Sains Malaysia Her research interests include preschool education, primary school education, and curriculum studies.

Correspondence concerning this article should be addressed to Jennifer Bauer at Jennifer_bauer@student.uml.edu, or Qing Zhao at qing_zhao@student.uml.edu

Current Issues in Education
Mary Lou Fulton Teachers College • Arizona State University
PO Box 37100, Phoenix, AZ 85069, USA

Manuscript received: 09/04/13
Revisions received: 11/30/2013
Accepted: 12/02/2013

Current Issues in Education
Mary Lou Fulton Teachers College • Arizona State University
PO Box 37100, Phoenix, AZ 85069, USA

Volume 17, Number 1 February 04, 2014 ISSN 1099-839X