



Cross-Cultural Reading Comprehension Assessment in Malay and English as it Relates to the Dagostino-Carifio Model of Reading Comprehension

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

The review of existing literature suggests that few researchers have adopted cross-language comparisons to explore how cultural background affects the assessment of reading comprehension of students. In this present study, the researchers independently reviewed and rated all the items of two reading comprehension tests translated from Malay into English. The original tests, based upon Dagostino-Carifio's (1994) model of reading comprehension, were developed in Malaysia for the purpose of evaluating reading comprehension abilities of K-6 students in Malaysia. Results indicated that some information in the test material are culturally based and are lost in translation. Findings imply the possibility of employing comparable tests with valid translation to evaluate reading comprehension abilities of second language (L2) students.

Keywords: reading comprehension, cultural background, cognitive process

In a world of increasing globalization and cross-cultural communication, how differences in cultural background affect the assessment of reading comprehension has become of interest to many researchers and educators (Read & Rossen, 1982; Reynolds et al., 1981; Porat, 2004). More specifically, previous research has considered the effects of general background knowledge as determined by culture on reading comprehension (Steffenson et al., 1979; Reynolds et al., 1981). Other research has considered the effect of the structure of the reader's language (Porat, 2004; Rajabi, 2009), the reasoning processes often used in the culture (Goetz, 1979; Thorndike, 1917; Pritchard, 1990), and variations that may exist in the instruction and assessment in different cultures (Johnston, 1984; Garcia, 1991; Medina, 2010). However, the designs of these studies did not use controlled or direct cross-cultural or cross-language comparisons, nor were the language equivalencies of the assessment used ascertained, this point being the more important of the these two points and a contributing factor to the significance of the present study.

As the initial step in the area of cross-cultural comparative research on this topic, the focus in the present study was to see how aspects of cultural background bear on the assessment of reading comprehension of students in the primary grades (1-6) of school using two Malaysian based tests of comprehension (Hashim et al., 2006) that were based upon the Dagostino and Carifio (1994) model of reading comprehension (Test I for grade 1-3, Test II for grades 4-6). In English translations of these two tests, the items were independently rated by three expert judges according to the elements of the Dagostino and Carifio reading comprehension and understanding model to see if the independent classifications of the translated test items agreed with the classifications of the original ones done by the Malay test makers, thus validating qualitative design characteristics of the tests (a type of qualitative construct validity) in both Malay and English, as well as the process used to translate the tests from Malay to English. Validating translations of tests relative to the equivalencies of the softer qualitative psychometric properties of the test is not typically done in multi-

language versions of tests, thus making this study important and so relative to the meaning of scores obtained from the test in either language, or any comparisons done. Therefore, this study focused on this validation first by completing the independent classification process of correct answers, reading comprehension levels, and skills for each test item; next, by examining the nature of the agreements of the ratings by expert raters; and third, by considering how the ratings reflect concepts underlying the Dagostino-Carifio Model of reading comprehension and understanding. This process allowed us to address the following research questions:

- What is the inter-rater agreement when the judgments of reading comprehension levels and skills of all raters are analyzed as a group?
- How do the three raters' judgments of levels and skills for each test item compare with the Malaysian Table of Specifications?

With this sense of the purpose and the significance of the study, we turn to a description of the theoretical perspective that is the underpinning of this study – the Dagostino-Carifio Model of reading comprehension.

The Dagostino-Carifio Model of Reading Comprehension

The original research, using the Malay version of the reading comprehension tests used in the present study, was based upon the Dagostino-Carifio Model of reading comprehension. This model is described here, and then discussed in relationship to the present study.

Core Principle of the Model

The Dagostino-Carifio Model of reading comprehension is organized around the principle that the process of message evaluation is integral to reading comprehension, and that it is used constantly throughout the reader/text interaction. This view that message evaluation occurs constantly and continuously is a general tenet of the cognitive view of information-processing and learning underlying what we say here (see Carifio, 2005). It permeates the concept of literacy that underlies our concept of reading comprehension. This idea of constant interaction contrasts with a strict sequential, hierarchical view of reading comprehension, which suggests a direct progression from literal to inferential/interpretive comprehension, and then to message evaluation of a text, followed by appreciation and application in well-known step-like depictions of “levels of comprehension” (e.g., Bloom et al., 1956; Johnson & Pearson, 1968; Pearson & Fielding, 1991; Anderson & Krathwohl, 2001). The Dagostino-Carifio Model of reading comprehension represents the synthesis of hierarchy and interaction, or skills and schema. In the case of a skills view, the

subskills of evaluation are specified clearly in lists of skills and levels of comprehension. In a schema view, reading evaluatively may be a more general process where the analytical skills and metacognitive strategies have been assimilated and synthesized to function as part of a less explicit evaluation process. The Dagostino-Carifio Model is an attempt to synthesize these two views of the reading process in relationship to reading comprehension and evaluation of text. What also is key to the discussion is that while more recent discussion have shifted focus to sociocultural perspectives, our discussions focus on the mind of the individual reader (McVee, Dunsmann, & Gavaleck, 2005).

The visual that follows is a representation of the model of reading comprehension (see Figure 1).

Key Terms. The following definitions of terms and subsequent text should help you to understand the Dagostino-Carifio Model of reading comprehension.

Message Evaluation. *Message Evaluation* of a text is part of a reader's response to what s/he exhibits in an interpretation of the text's message. The expectations for a text may influence a reader's response, and in turn, bias the message interpretation and message evaluation of the text. Further, an interpretation of the meaning may bias message evaluation. Thus, there is a dynamic nature to these processes. The reader's evaluation of a text also may vary with 1) the degree of objectivity and distance maintained, 2) the stringency of the application of criteria, and 3) the latitude of the selected criteria. This variation in the depth of message evaluation may be attributed to a reader's purpose in reading (Narveez, VanBroeck, & Rutz, 1999). Also message evaluation skills may vary as function of cognitive development as well as learning and experience. These terms are described here as they apply to the Dagostino-Carifio Model of reading comprehension and the Malay tests. In general, *Message Evaluation* requires the reader to apply criteria to the text, and make judgments about the text and the message of the text. *Message Evaluation* most closely describes the critical/creative level of comprehension used in the Malay tests.

Message Interpretation. *Message Interpretation* of a text is an intermediate response in the process of message evaluation of what a reader comprehends in a text. Message interpretation of the text may take at least two directions. First, when there is a mix of explicitness of text and closed, convergent thinking, message interpretation becomes closed, text-bound, perhaps even literal in nature. Second, a mix of implicitness of text and open, divergent thinking may lead to multiple message interpretations that are speculative in nature (Rosenblatt, 1978). The two important parts of message interpretation are closed, convergent thinking

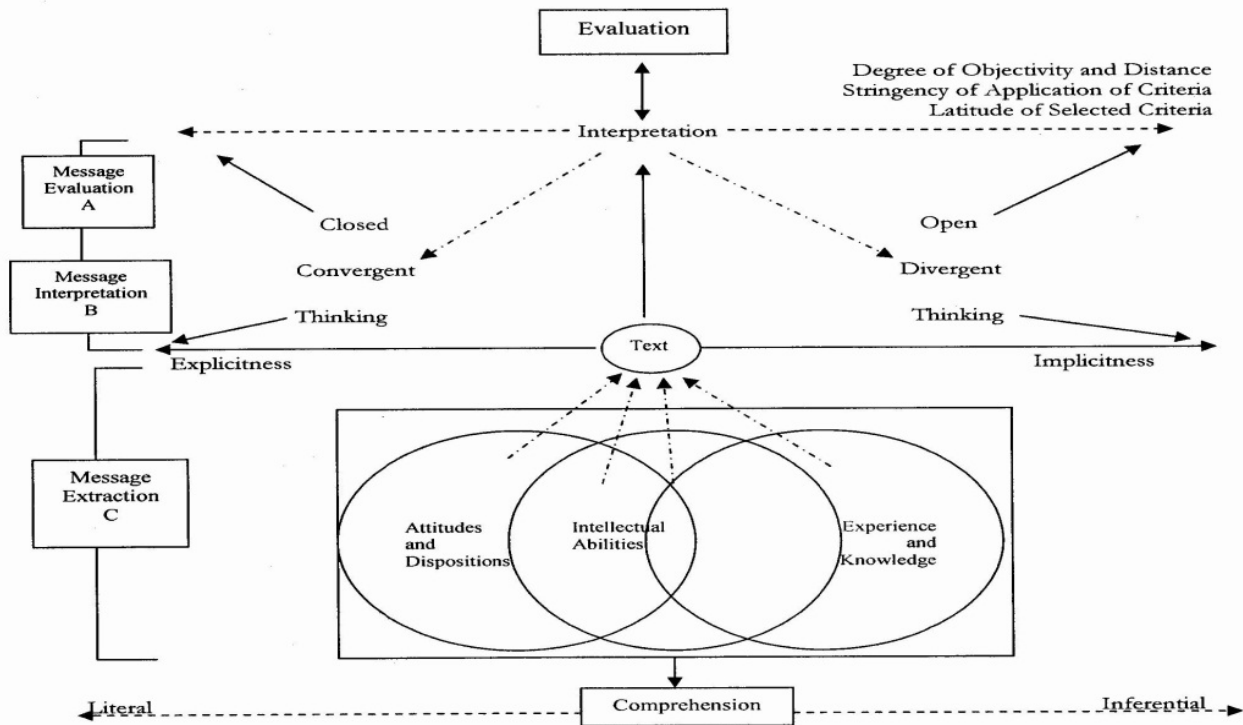


Figure 1. The Dagostino-Carifio Model of Reading Comprehension.

and open, divergent thinking. Both of these parts of message interpretation are described next even though this study was focused only upon closed, convergent thinking.

Closed, convergent thinking. Closed, convergent thinking is a style of processing text that limits the direction the reader takes in drawing conclusions. This pattern of thinking and processing usually leads the reader to a logical focused and limited set of conclusions or predictions about the meaning, interpretation, and evaluation of the text. The term closed suggests certain expectations and specified criteria that may influence thinking in a particular way.

Open, divergent thinking. Open, divergent thinking is a style of processing text that multiplies the directions the reader takes in drawing conclusions or establishing explanations and interpretations of a text. The terms open and divergent suggest the possibility of other than conventional reasoning or structuring of the details in developing a synthesis of information or ideas. Open and divergent do not rule out systematic thinking. However, interpretation may be other than generally expected, and perhaps more creative in nature where creative means a novel structuring of information or novel explanation of events. A reader's thinking may be internally consistent, but logically unexpected.

In general, *Message Interpretation* moves beyond the literal reading of a text to making inferences

of various types. It begins to use reasoning in the form of convergent, closed thinking to ascertain the meaning of a text, and it also may require applying what is understood of the message. Message Interpretation most closely describes the inferential level of comprehension used for the Malay tests.

Message Extraction. In general, *Message extraction* is simply getting the literal meaning from the text, with little or no inferencing/interpreting involved. It is the most basic level of comprehension in most representations of reading comprehension. Message Extraction is most closely described by the literal level of comprehension used in the Malay tests.

Text. Text represents the author's message through degrees of explicitness and implicitness. The degree to which the text is explicit or implicit influences how the reader comprehends and interprets the message of the text.

Reader. The experience with reading, in general, or with a specific text, style, or form, combined with the reader's knowledge of the topic and his ability to reason appropriately, varies with each reader and across readers (Anderson, Reynolds, Schallert, & Goetz, 1977). A trio of characteristics (knowledge and experience, intellectual abilities, and attitudes and dispositions) influences where the reader enters the literal/interpretive scale of comprehension. Maturity is the fourth variable, which encompasses the other three attributes.

Reading comprehension. The process of reading comprehension demonstrated by the reader moves along a continuum of literal to inferential/interpretive understanding in a manner that parallels the explicitness or implicitness of the text's message. What the reader understands of the text at this level is constructed by the reader and also provides the basis for message interpretation and message evaluation. If elements of the text are not comprehended, they may not be included in either the open, divergent thinking or the closed, convergent style of processing text. In turn, both message interpretation and message evaluation fluctuate because some of the information was not processed on one occasion but is on another. Learning and experience, among other things, can lead to the inclusion or exclusion of information from processing. In general, it should be noted that Message Extraction is the main component for the concept of Comprehension illustrated in the model and Message Interpretation and Message Evaluation are the main components for the concept of overall Evaluation illustrated in the model.

The Relationships of the Components of the Model of Reading Comprehension

With reference to Figure 1, we can think of message extraction, message interpretation, and message evaluation as horizontally shifting components of the model that move independently and interdependently. Message extraction covers the territory of reading comprehension from literal comprehension and recall to the beginning of inferential/interpretive processes based upon direct experience, general and specific knowledge, and reasoning. Although never completely separate from reader background, message extraction is more closely related to the cues in the text that carry the message and shape the presentation as developed by the author to be apprehended by the individual reader.

The trio of circles and the surrounding rectangle represent the reader. Primarily, the trio and the rectangle represent the interaction of knowledge and experience, attitudes and dispositions, intellectual abilities, and maturity. These interactions influence literal comprehension, inferences/interpretation and evaluation. Message interpretation and message evaluation draw on and act on message extraction. That is, they rely heavily on the reader's information and thinking, perhaps schema; yet, there is still a dependence on the text message and the way the reader's schema creates and shapes its interpretation of the text (Rosenblatt, 1978).

Each component has different goals. Message extraction depicts comprehending the message and its clearly implied conclusion. Message interpretation and message evaluation depict evaluating, with specific criteria, the message of the text. Message interpretation and message evaluation are the response part of the model where the response may be personal, subjective, or

criteria objective, given the reader's purpose and the criticalness of the judgments to be made.

When we look carefully at the reader in relationship to the process depicted here, we see the reader's ability to evaluate shifts dramatically depending on the subject matter or the form of the text. This feature characterizes it as a dynamic, process model. It is hoped that the mature reader will develop sufficient skill to raise general evaluative questions so that some sense of veracity of the text can be determined, regardless of the subject matter. However, this model implies that as the reader's knowledge base and thinking skills equal or surpass that of the text, message evaluation with specific criteria operates constantly so that the criteria can operate both before and after inferences are made. The criteria operate while the reader selectively and judiciously chooses and evaluates what is stated directly and implicitly derived from the text. The occurrence of this interaction means that an evaluation schema also may shape interpretation. Message Interpretation plays an important part in this model because it is here that we see message evaluation shaping comprehension. Here is where the reader's background selects relevant details for his or her purpose, and then constructs a whole that is meaningful. This meaning, or interpretation, may or may not be justifiable in light of the text's consistency of structure or the intent of the author. However, it is this meaning derived from the text, which the reader responds to evaluatively.

Finally, a shifting horizontally of the message extraction and message interpretation and message evaluation components depicts the uneven, dynamic nature of the comprehension/evaluation interaction. It is here that we see that as the reader has different purposes and goals in mind, he or she may elect a different set of criteria for message evaluation and, depending on the criticalness of the decision, that is, what is at stake, apply the criteria stringently, or leniently. In the model of reading comprehension, both message evaluation and message interpretation are part of our response to the text, perhaps one following the principles of logic and one following the aesthetics of language and themes. When both the mindset for message evaluation and the mindset for message interpretation are working simultaneously, the reader is able to respond to thought structure and style simultaneously and evaluate if the language is used effectively to sustain the thinking and the emotion of the writer. Both message evaluation and message interpretation are central to reading evaluatively as they relate to the variety of literate worlds envisioned by this model and the present work. In this model of reading comprehension, readers must become aware of themselves as readers so they do not simply accept the text. For readers to achieve a satisfactory evaluation of the text, they must cultivate an intellectual style that prompts

them to be self-conscious about the assumptions and the goals of their reading as well as to read deliberately from a different perspective.

The Model's Relationship to the Study

The Dagostino-Carifio Model of reading comprehension was the primary theoretical construct for the development of the Malay test, and in turn, should be the basis for the translation of the English version. It is the basis for the test for its view on having message evaluation as the main goal of higher-level reading comprehension. It also frames the levels of comprehension construct that shapes the levels of comprehension used in the Malay test. While the model does not identify specific skills explicitly, it acknowledges the skills view of reading and the legitimacy of identifying such skills, particularly as part of the reasoning process of reading. What it does propose with regards to skills as well as to levels is that they may not be represented in a strict sequential and hierarchical fashion, but that they may be more fluid in nature. What a classification process of a translation process examines is how fluid and dynamic the relationships among skills and between levels actually may be. So with this understanding of how the comprehension levels and skills transfer to the model, we move on to the present study.

The Study

Parameters and Limitations of the Study

The present work applies to the primary (grades 1-6) reader who may be approaching the early stages of formal reasoning rather than to the expert (grades 9-adult) reader who may be into the stage of formal reasoning. The materials considered in this study are non-technical, such as essays, fiction, poetry, journalistic writing, rather than technical, such as scientific or mathematical. The nature of cognitive thinking considered is convergent, critical thinking rather than divergent, creative thinking.

The Malay Test

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. The following section of this paper 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 as a theoretical basis, the 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 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. As a result of this work, the researchers established the following Table of Specifications (Table 1), which outlines the relationships between the comprehension levels and skills underlying the construct of both tests.

Defining the Comprehension Levels and the Comprehension Skills. The comprehension levels and the skills define the difficulty and the nature of the reading texts and the test items. In the present study, the comprehension levels of the test correspond to message extraction (literal), message interpretation (inferential), and message evaluation (critical/creative) of the Dagostino-Carifio Model of reading comprehension.

Comprehension levels. (a) Literal (message extraction) – refers to memorization of facts in texts where information is explicitly stated at a basic level of thinking; (b) inferential (message interpretation) – refers to the ability of pupils interpreting meaning where they need to use overt information with intuition and experience requiring a high level of thinking assessed by the Malay tests; (c) critical/creative (message evaluation) – refers to ability to do an overall evaluation towards a certain information or idea which is read, precision or suitability of the given information of a new idea; needs divergent thinking outside literal/inferential depends upon knowledge and personal experience of the pupils, but focuses on convergent critical thinking.

Comprehension skills. There are ten comprehension skills that are assessed by the Malay tests: (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 comprehension to what is called deep or deeper understanding, which is a first step towards what is called evaluative reading.

Selection and Development of Reading Texts

Types and content of the texts. There are several types of text that make the test broad in scope and representative of various types of reading of non-technical materials that are encountered in daily reading situations. 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

Table 1
Table of Specifications for Malay Reading Comprehension Tests

Comprehension Category	Code	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

texts are 100 words or 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 comprehension than those used for grades 4-6. A research group, three 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 items. The questions and answers 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. Each test consists of 50 multiple choice items designed to evaluate reading comprehension with consideration given to skill ability and comprehension level. Some specific things were considered in the item development. They are as follows: a) arrangement of each item was based upon 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 Table of Specifications includes the classification by reading comprehension level and skill for each test item. Both Malay tests were built from the same Table of Specifications (Table 2).

Design and Choice of Answers and Distractors.

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 distractors 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.

Reliability Measures of the Two 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

Cross-Cultural Reading Comprehension Assessment in Malay and English as it Relates to the Dagostino-Cariffo Model of Reading Comprehension

Table 2

Malay Table of Specifications Including Test Items by Classification, with this General Template Being the Same for Test I and Test II

Comprehension Category	Code	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	
			21	
Inferential (F)	F1	interpreting main idea	25	
			29	
			43	
			17	
	F2	interpreting important point	26	
			48	
			18	
	F3	interpreting comparison	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	

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 would that are currently being done.

Methodology

Procedures

The Malay versions of the tests were translated into English, and three expert raters completed the inter-rater judgments of answers, levels and skills. The next section of this paper describes this translation process as well as the raters' process for

classifying test items by answer, level and skill.

The Translation Process

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 have been 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 of Massachusetts Lowell. 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.

Procedures for the Inter-rater Judgments of the English Version of the Two Tests

The raters have either Ph.D. in language arts and literacy, or were completing work for that degree. One of the raters speaks both English and Chinese, and another works with young children from several cultures and language backgrounds. The three raters began their evaluation of the English tests by first meeting to discuss the relationship of the components of the Dagostino-Carifio Model of Reading Comprehension and the two Malay tests, as well as the procedure for rating each test item with regards to answers, comprehension, and skills. The three expert raters completed the inter-rater agreements by first reading each item of Test I and Test II independently, and then answering each test item by selecting what they thought was the correct answer from the four choices provided. Next, each rater then classified individual test items by level of comprehension that they were testing. The levels of comprehension were 1) literal, 2) inferential/interpretive, and 3) critical/creative, a category comparable to evaluation in the Dagostino-Carifio Model of reading comprehension and understanding. Finally each rater classified individual test items by reading comprehension skills that they were testing. The skills were classified using to the Malay Table of Specifications (i.e. main points, word meanings, etc.).

After the independent readings and ratings were

completed, the raters compared their judgments for all three types of ratings – answers, comprehension levels, and skills. 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 item ratings. The test content was studied in relationship to the Dagostino-Carifio Model of reading comprehension to see how the tests reflect the concepts that are central to the model. Finally, there was further discussion on the influences of translating texts from one language to another on the ability to rate the items with regards to cultural differences in the test such as references to place, customs or language use, as well as how the tests may be interpreted by any individuals taking these tests.

Results and Data Analysis

The results and data analysis section of this paper presents the data and its interpretation for two questions. These questions are:

- What is the inter-rater agreement when the judgments of reading comprehension levels and skills of all raters are analyzed as a group?
- How do the three raters’ judgments of levels and skills for each test item compare with the Malaysian Table of Specifications?

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 accuracy of the translation of both tests, each rater completed the tests individually, and then compared their answers. The agreement rate was 98% for both tests ($r > .98$), indicating that the translation did not affect the meaning of the test. With the stability of the translation confirmed, the raters individually used the Malay Table of Specifications to assign a comprehension level and skill to each test item of both Test I and Test II. Once this work was complete, the raters gathered to compile the quantitative data and examine the results.

The first question addressed was, “What is the inter-rater agreement when the judgments of reading comprehension levels and skills of all raters are analyzed as a group?” Table 3 presents a comprehensive look at the frequencies and percentages of rater agreements for both Test I and Test II, further broken down by levels and skills. The square roots of the agreement percentages approximate the inter-rater correlation coefficients.

As can be seen from Table 3, agreement between the raters was very high in regard to their classifications of test items by comprehension level and skill ($r > .91$).

Table 3
Agreement of Inter-rater Classification of Test Items by Comprehension Level and Skill

Test I: Rater Agreements				
Percentages of Rater Agreements by Classification of Comprehension Levels				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed on level classification	48	.96	96	.98
2. Raters disagreed on level classification	2	.04	100	
	50	100%		
Percentages of Rater Agreements by Classification of Skills				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed on skill classification	41	.82	82	.91
2. Raters disagreed on skill classification	9	.18	100	
	50	100%		
Test II: Rater Agreements				
Percentages of Rater Agreements by Classification of Comprehension Levels				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed on level classification	49	.98	98	.99
2. Raters disagreed on level classification	1	.02	100	
	50	100%		
Percentages of Rater Agreements by Classification of Skills				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed on skill classification	47	.94	94	.97
2. Raters disagreed on skill classification	3	.06	100	
	50	100%		

Table 4

Comparison of Raters' Classification of Test Items by Comprehension Level and Skill with the Malaysian Table of Specifications for Test I and Test II

Test I: Raters' Classification Compared to Malay Classification of Test Items				
Percentages of Agreement by Classification of Comprehension Levels				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed with Malay classification	33	.66	66	.81
2. Raters disagreed with Malay classification	17	.34	100	
	50	100%		
Percentages of Agreement by Classification of Skills				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed with Malay classification	27	.54	54	.73
2. Raters disagreed with Malay classification	23	.46	100	
	50	100%		
Test II: Raters' Classification Compared to Malay Classification of Test Items				
Percentages of Agreement by Classification of Comprehension Levels				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed with Malay classification	41	.82	82	.90
2. Raters disagreed with Malay classification	9	.18	100	
	50	100%		
Percentages of Agreement by Classification of Skills				
Type of Agreement	Frequency	Percent	Cum. Percent	R
1. Raters agreed with Malay classification	44	.88	88	.94
2. Raters disagreed with Malay classification	6	.12	100	
	50	100%		

After completing this analysis, we next asked, "How do the three raters' judgments of levels and skills for each test item compare with the Malaysian Table of Specifications?" Table 4 presents a comprehensive look at the frequencies and percentages of rater agreement with the Malaysian Table of Specifications for both Test I and Test II, further broken down by levels and skills. The square roots of the agreement percentages approximate the correlation coefficients.

As can be seen from Table 4, for Test I, the raters assigned 66% of the items the same comprehension level as the Malay Table of Specifications; for Test II, the agreement rose to 82%. When evaluating skills, the raters assigned 54% of the items the same comprehension skill as the Malay Table of Specifications for Test I; for Test II, the agreement rose to 88%. What these results indicate is that the classification of the difficulty of a task is very different than finding the correct answer to a particular question.

It is possible that the classification of the skill and level of each individual test answer is not as easily done as merely determining the answer itself because these things are not as discrete as one would think. One's background and cultural knowledge most certainly could affect how he or she interprets a question, and therefore

how he or she would classify what the question is asking the test-taker to do in terms of skill.

**Conclusions, Discussion and Implications
Summary of Raters' Discussion Regarding
Relationship between Test Content and Model**

One issue that came to light while discussing the results was that the classification of skills is not as discrete as we would think, or would like them to be. What skills are literal versus evaluative are often differentiated by a fine line. For example, reading for the "main idea," is dependent on the question and the reading passage, which could be asking for a literal interpretation as ascertained by the test, or an inferential interpretation based on the presumed prior knowledge of the reader.

Furthermore, another concern was that as raters we are not primary or middle school students, therefore we may have a greater store of prior knowledge or reading schemas than the students for whom the test is geared. We raters then applied this prior knowledge when selecting our answers for each test question. As a result, we may interpret a question differently than the students taking the test, which results in a greater number of discrepancies between the judgments of the raters and the Malaysian answer key. The Dagostino-Carifio Model supports this finding. The interaction of knowledge and

experience, attitudes and dispositions, intellectual abilities, and maturity affects how a reader evaluates a text; therefore adults read a test that is geared toward primary school students differently.

It might be hard assessing the evaluative skills of the primary school reader because they are so interwoven with cognitive thinking abilities and dependent on knowledge beyond just basic literal reading skills. We might not see this particular type of interconnectedness in the test because test is focused on the primary school reader and the model is geared more toward the expert reader and higher thinking skills. Additionally, we are dealing primarily with non-technical texts of literature, rather than text of a scientific or math nature. At the heart of the Dagostino-Carifio Model of reading comprehension is a focus on evaluation, a skill not usually mastered by the primary school reader.

Summary of Raters' Discussion Regarding Translation Concerns

A concern shared by the raters was the prospect that one's level of knowledge could possibly affect his or her understanding of a translated version of the test. The specific nature of background knowledge that is entrenched in culture can affect the test taker's response. For example, in Test I, the reading passage used for questions 21-24 is a letter to a father from his son. In the letter, the son describes a trip he took approximately one week after Independence Day. Question 22 then asks, "The visit was held in _____?" and gives four options: a. June, b. July, c. August, and d. September. An American student taking the test would answer "b," as Independence Day in the United States is celebrated in July. However, a Malaysian student would answer "c," as the Malaysian Independence Day is in August. This is a clear example of how cultural differences do not necessarily translate appropriately. The events, activities, objects and timeframes referred to in the test material are very culturally based and end up lost in translation. However, this point clearly suggests that the details of multiple language versions of many tests should be tailored and customized to the culture in which the test is going to be used and particularly if international comparisons are going to be made.

Future Implications

One question that arose during the discussion was, "Could this translation process be applicable to many tests across many languages?" If the translation process were valid, it would allow us to make reliable cross-cultural comparison because you could get past issues of culture and prior knowledge. In this world of increasing globalization, it is important for us to explore how people can handle a second language because you can then give them more comparable tests if you have detailed answers on this particular point. By applying a different measure for comprehension level, perhaps one that looks at cognitive levels of understanding rather than

the more traditional measures of reading comprehension, it may be possible to more precisely classify the comprehension level and skill of each test item. One such measure or framework would be the most recent version of Bloom's revised taxonomy. Both of these avenues could be explored in future studies.

References

- Anderson, R., Reynolds, R., Schallert, D., & Goetz, E. (1977). Frameworks for comprehending discourse. *American Educational Research Journal*, 14, 367-381.
- Anderson, L. E., & Karthwohl, D. (Eds.). (2001). *A Taxonomy for Learning, Teaching and Assessment*. New York: Longman.
- Bloom, B. S., Englehart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I. Cognitive domain*. New York: McKay.
- Carifio, J. (2005). An integrated information processing theory of learning. *Proceedings of the Eighth International History, Philosophy & Science Teaching Conference, Leeds, England*, 101-173. Retrieved from <http://www.ihpst2005.leeds.ac.uk/papers/Carifio.pdf>.
- Dagostino, L., & Carifio, J. (1994). *Evaluative Reading and Literacy: A Cognitive View*. Boston: Allyn and Bacon.
- Garcia, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, 26(4), 371-392.
- Goetz, E. T. (1979). Inferring from text: Some factors influencing which inferences will be made. *Discourse Processes*, 2, 179-195.
- Hashim, N. H., Yahaya, A. S., Aziz, A. R. A., Mohammed, A. R., Othman, H., Ramly, I.,...Lah, Y. C. (2006). *Reading comprehension in Malaysia primary grades 1-6*. Penang, Malaysia: Universiti Sains Malaysia.
- Johnston, P. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, 19(2), 219-239.
- McVee, M. B., Dunsmore, K., & Gavalek, J. R. (2005). Schema theory revisited. *Review of Educational Research*, 75(4), 531-566.
- Medina, C. (2010). "Reading across communities" in biliteracy practices: Examining translocal discourses and flows in literature discussions. *Reading Research Quarterly*, 25(4), 273-295.
- Narvaez, D., vanden Broeck, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, 91(3), 488-496.
- Pearson, D., & Fielding, L. (1991). Comprehension instruction. In Barr, Kamil, Mosenthal, and

- Pearson (Eds.), *Handbook of Reading Research*, (815-860). White Plains, New York: Longman.
- Porat, D. A. (2004). It's not written here, but this is what happened: students' cultural comprehension of textbook narratives on the Israeli-Arab conflict. *American Educational Research Journal*, 41(4), 963-996.
- Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25(4), 273-295.
- Rajabi, P. (2009). Cultural orientation and reading comprehension models: the case of Iranian rural and urban students. *Novitas-ROYAL*, 3(1), 75-82.
- Read, S. J., & Rossen, M. B. (1982). Rewriting history: The biasing effects of beliefs on memory. *Journal of Social Cognition*, 1(3), 240-255.
- Reynolds, R. C., Taylor, M. A., Steffensen, M. S., Shirey, L. L., & Anderson, R. C. (1981). *Cultural schemata and reading comprehension*. (Tech Report No. 201). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Document Reproduction Service No. ED 201 991).
- Rosenblatt, L. (1978). *The reader, the text and the poem: The transactional theory of literary work*. Carbondale, Ill.: Southern Illinois University Press.
- Steffensen, M. S., Joag-dev, C., & Anderson, R. C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 15, 10-29.
- Thorndike, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology*, 8, 323-332.

Article Citation

Dagostino, L., Carifio, J., Bauer, J. D. C., & Zhao, Q. (2013). Cross-cultural reading comprehension assessment in Malay and English as it relates to the Dagostino-Carifio model of reading comprehension. *Current Issues in Education, 16*(1). Retrieved from <http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1063>

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
50 Father Morissette Boulevard, Lowell, MA 01852
jennifer_bauer@student.uml.edu

For the past eleven years, Jennifer D. C. Bauer has taught digital media and visual art at Lowell High School, an inner-city public school in Massachusetts. 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 began her graduate study at Graduate School of Education, University of Massachusetts Lowell in 2006. She is currently a doctoral candidate and a research assistant in the Language Arts and Literacy program. 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.



Current Issues in Education

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

Manuscript received: 8/26/12

Revisions received: 1/8/13

Accepted: 1/12/13



Current Issues in Education

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

Volume 16, Number 1

February 14, 2013

ISSN 1099-839X

Authors hold the copyright to articles published in *Current Issues in Education*. Requests to reprint *CIE* articles in other journals should be addressed to the author. Reprints should credit *CIE* as the original publisher and include the URL of the *CIE* publication. Permission is hereby granted to copy any article, provided *CIE* is credited and copies are not sold.



Editorial Team

Executive Editors

Melinda A. Hollis
Rory Schmitt

Assistant Executive Editors

Laura Busby
Elizabeth Reyes

Layout Editors

Bonnie Mazza
Elizabeth Reyes

Recruitment Editor

Hillary Andrelchik

Copy Editors/Proofreaders

Lucinda Watson

Authentications Editor

Lisa Lacy

Technical Consultant

Andrew J. Thomas

Section Editors

Hillary Andrelchik
Michelle Crowley
Ayfer Gokalp
Darlene Michelle Gonzales

Courtney Hart
David Isaac Hernandez-Saca
Sultan Kilinc
Yoonsu Kim

Linda S. Krecker
Carol Masser
Bonnie Mazza
Constantin Schreiber

Faculty Advisors

Dr. Gustavo E. Fischman
Dr. Jeanne M. Powers
