The Translation and Translation Verification of the PIRLS Reading Questionnaires for

Saudi Students

By

Fahad Abdullah Alharbi
B.S., Taibah University
M.S.Ed., The University of Kansas
P.hD., The University of Kansas

Submitted to the graduate degree program in the department of Educational Psychology in Education and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science in Education.

_____

Dr. John Poggio, Chairperson
Department of Educational Psychology

_____

Dr. Jonathan Templin
Department of Educational Psychology

_____

Dr. Vicki Peyton
Department of Educational Psychology

Date Defended: 04/22/2017

The Thesis Committee for Fahad Alharbi

certifies that this is the approved version of the following thesis:

THE TRANSLATION AND TRANSLATION VERIFICATION OF THE PIRLS
READING QUESTIONNAIRES FOR SAUDI STUDENTS

_____

Dr. John Poggio, Chairperson

Date approved: 04/22/2017

ABSTRACT

When translating psychological measures from their original language to another language, after translation an assumption is made that the measurement assesses the same construct(s) in the original language and for those in the group of the translated language. If this assumption is not met, a translation problem occured and then the measurements are not comparable across cultures because they are not assessing the same variables. This study investigated the quality and accuracy of select variable translations of the Progress in International Reading Literacy Study (PIRLS) for Saudi Arabian students. This study only focused on the quality of translation of three reading affective constructs (reading motivation, reading attitude, and reading confidence). In their translation and verification procedures, the IEA used the Expert Translation Method (ETM, Mullis et al., 2009); they did not use a common method (given citations here) known as Backward Translation Method (BTM, from English to Arabic and then from Arabic to English) as a step of translation validity from English to the Arabic language. This investigation conducted the Backward Translation Method as a step of validation to evaluate whether the final Arabic PIRLS affective scales version were the same for those three constructs. After comparing between the IEA ETM translation version and the BM translation version, the researcher found that certain items were translated by the IEA ETM for Saudi students were above their level of reading understanding and thus comprehension while other items were not semantically equivalent. Results of this study advise researchers to proceed with caution as some attitudinal affective items are not comparable across the two cultures and not within the Saudi culture.

ACKNOWLEDGEMENTS

I am grateful to Allah for the good health and wellbeing that were necessary to complete this thesis.

I would like to express my sincere gratitude to my advisor and theses chair, Dr. John Poggio, for his continuous support of my master study and related research and for his patience, motivation, and immense knowledge. His guidance helped me during the research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master study.

Besides my advisor, I would like to thank the rest of my master committee, Dr. Vicki Peyton and Dr. Jonathan Templin, for their insightful comments and encouragement, which incentivized me to widen my research to include various perspectives.

My deepest appreciation is acknowledged to my wife, Fadwa Alharbi, who supports me continuously in my studies and my life. Also, I offer appreciation to my father and mother for their support, encouragement, and persistent prayers. Finally, I am grateful to all the other people who supported me directly or indirectly as I worked to create this research.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

Over the past two decades, the use of international assessments of educational

achievement has increased. To illustrate, the International Educational Achievement (IEA),

which conducts the Progress in Reading Literacy Study (PIRLS), and the Trends in International

Mathematics and Science Study (TIMSS) assessments have grown from 41countries in 1993 to

69 countries in 2011(L. Rutkowski, Gonzalez, Joncas, & von Davier, 2010)

The PIRLS assessment instruments were developed originally in English, which the

participating countries then translated and adapt into their country's languages of instruction.

PIRLS seek a high quality of translation, so they had multiple rounds of reviews by linguistic

and assessment experts to make sure that the original version, again, written in English, is

equivalent to the national language(s) for each participating country. By doing a process of

translations and verifications, the IEA was seeking for high quality translations to make the

instruments adapt to each country's context and education system by using a testing system that

was comparable across countries (Mullis et al., 2009). According to Mullis, Martin, Kennedy,

Trong, and Sainsbury (2009) the translations and verification include

- Student achievement test items (blocks of items for TIMSS and reading passages and items for

  PIRLS and pre PIRLS)

- Background questionnaires for students, parents (PIRLS and pre PIRLS only), teachers, and

  school principals

- Covers and directions (for each achievement booklet and background questionnaire)

- Online covers and directions (for teacher and school questionnaires and online data collection

  only)

As explained in Mullis et al.,(2009) the process of IEA's translation and translation

verification, know as the Expert Translation Method (ETM) occurs two times. The first time

occurs during the field test and the second time occurs for the assessment itself. There are 69 countries participants who worked on preparing 215 sets of translated achievement tests and 170 sets of translated background questionnaires for both TIMSS and PIRLS at the fourth and the eighth grade levels. One of the critical challenges that IEA faced is there are 58 languages. Therefore, the IEA needs to make sure that the original instrument is equivalent to all languages that are used by a country. English was the most common for the PIRLS assessment (16 countries) followed by Arabic (7 countries). For PIRLS, 17 countries administrated the test and the questionnaires in more than one language. For example, Saudi Arabia administrated the achievement test and questionnaire in both English and Arabic. Starting in 2007, the TIMSS and PIRLS international center developed the TIMSS tests and questionnaire in an Arabic language version. According to the IEA center, "the translation was developed through an extensively collaborative process between teams of expert translators and reviewers familiar with the terminology used in specific school subjects in different Arabic-speaking countries" (Mullis et al., 2009, p.10). The center wanted to make sure all facets of the test translation was accurate across the different counties in the Arabic world. To do that, the IEA cooperated with an independent translation agency, Brantra in Belgium (Mullis et al., 2009).

The Brantra process as contracted by IEA involves a group of skilled translators from different Arabic-speaking country who develop an initial translation of the international version of PIRLS 2011 for the fourth grade field test instruments, which included student achievement tests and questionnaires for students, and questionnaires for teachers, and school principals. After that, experts with collective experiences in the specific school subjects for the fourth grade reviewed and revised the translation "paying particular attention to the conformity of terminology with usage in school textbooks in a variety of Arabic-speaking countries" (Mullis et al., 2009, p. 12). Those experts created comprehensive notes, including any translation or

adaptation issues that might necessitate adaptation for the national context. After the field test was conducted, the same team of translators and reviewers reviewed and updated the Arabic version with a list of changes based on the international, i.e., the English version. However, in their manual IEA did not mention if the field test was conducted for each Arabic country or if it was just a sample of the Arabic world without considering across the Arabic world (22 countries) cultural differences (Mullis et al., 2009).

After developing the initial Arabic version for the PIRLS, the instrument was given to each participating country that was allowed to have a skilled translator translate the achievement test and the questionnaire based on the directions provided by the TIMSS and PIRLS International Study Center. It wanted to make sure the national versions of the instruments were consistent with the international version, so each country was allowed to make national adaptations as necessary. In addition, each country was allowed one or multiple independent translators make sure the translations were appropriate for all students. It was recommended that each country prepare translations in more than one language including professionals proficient in both English and Arabic languages to make sure that the translations were equivalent across languages. In fact, the IEA center required that each country hire high quality translators and reviewers. In addition, they list the following essential qualifications for translators that the country must follow:

- Excellent knowledge of English

- Excellent knowledge of the target language

- Experience in the country's cultural context

- Experience translating texts in the subject areas related to the TIMSS and PIRLS assessments (mathematics, science, and literary texts, respectively).

Also, the reviewers were expected to have enough experience in the fourth or eighth grade for the target population and excellent knowledge of both languages and the country's cultural context in order to assess the translation's readability and accuracy for that specific population. According to IEA, the translators and reviewers were asked to ensure that the following:

- The translation is at an appropriate level for the target population.

- No information is omitted, added, or clarified in the translated text.

- The translated text has the same meaning and uses equivalent terminology as the international version.

- Idiomatic expressions are translated appropriately, not necessarily word for word.

- The translated text uses correct grammar, punctuation, qualifiers, and modifiers as appropriate for the target language.

During the process of translating, each country must follow internationally agreed-upon standards for preparing instruments for their country's students. These procedures require each country to be responsible at both the national and the international level:

At the national level, countries are responsible for translating and/or adapting the international achievement tests and questionnaires according to TIMSS and PIRLS international guidelines, conducting an internal review of their appropriateness and quality, and documenting their adaptations for reference at later stages. Even for those countries whose survey language is English, adaptations are required to suit the variation of English used in the national context. (Mullis et al., 2009, p.15)

Also, the countries where Arabic language is used needed to confirm the specific national usage and context for their students and instruments. Next, at the international level, the IEA looked at the translation and verification for each country to make sure the translations met international

standards. Then, during the verifications, the National Research Coordinators (NRC) reviewed

the feedback from each country to revise their material as needed; they updated their

documentation for use during processing and analysis. Eventually, before assessment data

collection, the prior field test was conducted again to make sure the separate verification for

items was designed to measure trends from previous cycles (Mullis et al., 2009).

Finding appropriate terms and expression in the target language that have the same

meaning and style of text as the international version was the main challenge in translating the

PIRLS achievements test. It is important to make sure that changing or replacing a term by a

translator does not affect the meaning or the difficulty of an item. On the other hand, the

questionnaires have many terms and expressions that need to be adapted to be appropriate for the

education system and national context for the target population of each country. In the

questionnaires, countries can add a limited number of national interest items to the questionnaire.

All national questions must be approved by the TIMSS and PIRLS International Study Center

(Mullis et al., 2009).

Statement of Problem

When translating from one language to another language, there are three main issues that

should be considered. The first issue is the semantic equivalence, which means the words and the

structure of each item in the translated text has the same meaning and expression as the source

language. For example, when translating a construct from English to Arabic language, it is

important to make sure that the meaning of each item has the same meaning in both languages;

otherwise, that specific item will not measure the same thing. The second issue is conceptual

equivalence. When measuring a concept across countries, the concept needs to be measured at

the same level across groups even though the wording to describe it might be different. The third

issue is normative equivalence, which means the translation should be in high equivalence to address the social norms that might be different across cultures.

The IEA center purports to do a qualified and experienced job to make sure the translation has a high quality of translation across languages and met the standard of international version. We will refer to the IEA approach as the Expert Translation Method (ETM). Yet, Saudi Arabia had low scores in some constructs that are in students' booklets. For example, in the construct of reading motivation, reading attitude, and reading confidence for the fourth grade students, the Saudi group scored Cronbach's $\alpha = .7, .5$ and $.55$, respectively, whereas in the English home country these reliabilities were $\alpha= .84$, $\alpha= .76$, and $\alpha= .71$respectively. These low reliability indices for the Saudi population might be caused by the three issues that had been discussed previously regarding to the translation from English, to the original version of international to the Arabic version, the translated version. To further study this issue, this researcher will use the Backward Translation Method (BTM) (Mullis et al., 2009). to verify and evaluate the IEA ETM translation version for Saudi Arabia in the constructs (reading motivation, reading attitude, and reading confidence) (Mullis et al., 2009). The research will seek to answer the following questions:

1. Does the backward translation (BTM) differ from the IEA ETM translation for Saudi students' reading motivation scale in the PIRLS assessment?
2. Does the backward translation differ from the IEA translation for Saudi students' reading attitude scale in the PIRLS assessment?
3. Does the backward translation differ from the IEA translation for Saudi students' reading confidence scale in the PIRLS assessment?

Significance

Before using achievement test results to evaluate and compare across countries, it is

important to make sure that the measurement is comparable across cultures and native languages.

Confirming the accuracy of translation between the Saudi culture and American cultures on

those three constructs (reading motivation, reading attitude, and reading confidence) will help the

researchers to use those scales with greater confidence when making comparisons between Saudi

and United States students.

CHAPTER 2
Literature Review

The history of modern international assessment of student skills began in the early 1960s when the Evaluation of Educational Achievement (IEA) conducted the First Mathematics Study (FIMS) (D. Rutkowski, Rutkowski, & von Davier, 2013). In recent decades, international comparison studies have attracted attention, and many countries are participated in international assessment in reading, math, and science studies (Rindermann, 2007). TIMSS, PIRLS, and PISA are the most common international assessments in use around the world. The field of international assessment continues growing and includes assessments and surveys in other subjects and areas such as International Computer and Informational Literacy Study (ICLS), Civic Education Study (CIVED), Programme for International Assessment of Adult Competencies (PIAAC), and the Teaching and Learning International Survey (TALIS) (D. Rutkowski et al., 2013). Kamens and McNeely (2010) believe that a establishing and administrating international assessments will continue growing.

The main goal of the assessments organizations such as OECD, UNESCO, and PASEC is to assess educational achievement beyond national boundaries (Wagemaker, 2013). Several concerns have been raised regarding international surveys. One of those issues is to what extent these assessments have an impact on the education system of the participating countries (Wagemaker, 2013). In some developing countries, the policy makers are molding teaching practices after TIMSS, and they have begun to change the discourse on and diction of classroom instruction and curriculum (Elley, 2005).

International assessments have complex challenges, especially when the results are comparable across groups from different cultures(Glas & Jehangir, 2014). Therefore, it is important to ensure that framing items, which underlays the construct, are stable across cultures and countries. It is possible that in questionnaires, cultural bias may be more prominent:

First, it is no minor task to define constructs such as the socioeconomic status or the pedagogical climate in such a way that they allow for comparisons over countries and cultures and, second, culture-related response tendencies may bias the comparability between countries and cultures. (Glas & Jehangir, 2014, p. 98)

International educational assessments are administrated in many countries to inform decisions regarding educational policy, program, evaluation curriculum, and development. Scores across those countries are used to compared the students' achievement from one country to other countries (Cook, 2006). Comparability is required to create a scale to measure the same constructs across groups being compared with a similar level of uncertainty. It has been know that the use and the interpretation of the test score are the main issue for validity and measurement need to provide evidence to approve their validity across different countries (Sandilands, Oliveri, Zumbo, & Ercikan, 2013).

For PIRLS, students in the fourth grade were tested in their ability to understand written language forms required by society or valued by the individual. In addition, students could construct the meaning from a variety of texts (Mullis et al., 2009). According to Rindermann (2007), this type of task is similar to the reading task in other types of international measurement such as PISA and IEA reading studies. However, PIRLS were adapted to be easier with short tests, using common and less abstract words, and with less cognitive complex demand (Rindermann, 2007).

Working on cross-cultural research includes additional methodological challenges when compared to domestic research. If those additional methodologies are not addressed, the risk of inferential will be increased (Singh, 1995). Studies illustrate that constructs and concepts might require culture-specific attributes and meanings. Those should be explicitly taken into account when using and interpreting data across cultures (Church, 2010). One of those issues that need to

be considered when using data across cultures is the language of a questionnaire and its effect on how respondents answer the same question (Harzing, 2005). The researcher needs to make sure that those questions are equivalent in terms of their adopted constructs, measures, and samples (Mullen, 1995).

When the research involves a population that do not speak the original language that the survey was designed for (such as English for PIRLS), more attention and procedures should be undertaken in order to make sure that instrument is valid and reliable for the target population (Peña, 2007). In cross-cultural research, bias is a distinct threat to validity in translation of methods. Peña states that

An important principle for such a discussion is the notion of fairness in test development. Fairness is evaluated in the context of the goals or function of the test instrument. Definitions of fairness include equal treatment in context and purpose of testing, and comparable opportunity to demonstrate abilities on the construct the test is intended to measure. (2007, p.1256)

The principles of fairness and equity are to give people in different populations the opportunity to administrate the constructs equally. In fact, direct translation from language to language is not the only source of bias, but also instructions to the participants and the content of instruments might be sources of bias. When translating the questionnaire from one language to another, it is essential to not only consider the linguistic equivalence but also to consider the functional equivalence, cultural equivalence, and metric equivalence (Arnold & Matus, 2000; Peña, 2007; Rogers, Gierl, Tardif, Lin, & Rinaldi, 2003). Linguistic equivalence refers to translating a measurement from one language to another by using methods such as back translation or expert review (Hambleton, 2001). Functional equivalence refers to t the construct that measures the same target of behavior (Greenfield & Suzuki, 2006). Cultural equivalence refers to the way that the respondents to the items of the constructs have the same meaning across cultural linguistic

groups (Alonso et al., 1998). Metric equivalence referrers to the difficulties of each item across different groups; items are expected to be equal for both different groups, and if not, that means items are difficult to understand for one group compared to another (Kim, Han, & Phillips, 2003).

According to Beaton, Bombardier, Guillemin, and Ferraz (2002), it is not easy to achieve all types of equivalence when translating and asking questions about attitudes and opinions because the ideas might be more abstract, certain cultures do not discuss particular beliefs with strangers, and a concept might not be relevant throughout the world. Also, issues might be occur when a questionnaire is translated too closely, meaning the translation focuses on the words and not the meaning of the questions. In addition, when writing different questions than the original questions, an unidiomatic or improper use of the target language can be an unnecessarily complicated or awkward text will be created (Harkness, Pennell, & Schoua-Glusberg, 2004). There are two recognized methods of translations. The first method is called the expert method, where researchers rely on experts who know both languages Greenfield & Suzuki, 2006). The second method is called backward translation (Greenfield & Suzuki, 2006), where one or more translators translate the original language to a different language, and then another independent translator translates the newly translated form back to the original from. At that point, the new form will be compared to the original form to find out how they differ from each other and adjustments are made. Pan and De La Puente (2005) suggest five steps for translating from one language to another, which are: prepare, translate, pretest, revise, and document. They recommended having at least two translators perform translation; those translators should be experts in the subject matter and have knowledge in survey design and an adjudicator. Hurtado, Angeles, Blahut, and Hays (2005) indicated that the main roles for translators are to translate the text accurately, grammatically correct, written in an appropriate way for the population's reading

level, and sensitive to regional variations. The reviewers make sure the translators' text is accurate, written for the population level, and that all terminologies are correct and can be understood by most of the target population. In fact, the translators and reviewers should be experts in both languages and have excellent experience in translating documents within the field of the questionnaire. The researchers recommended different steps for both methods. In their research, Beaton et al. (2002) suggested six (6) steps for using the backward translation approach:

1. **Translation:** At this stage, a researcher will invite at least two independent forward translations. Both translators need to have excellent knowledge in both English and Arabic languages, have experience in the Saudi Arabian's cultural context, and have experience in translating texts in the area of reading. By having two bilingual translators, it will be possible to compare their versions to identify discrepancies indicative of ambiguous wording within the original measurement or other problems.

2. **Synthesis:** During synthesis stage, a third bilingual person with the same qualities of the other two translators reconciles a discussion between the two translators to develop one version of the measurement.

3. **Back Translation:** Another independent person, who is blind to the original measurement, will be invited to translate back the new Arabic version into the original language (the English version) and then compare it to the original document to check the validity of the translation

4. **Expert Committee Review:** The final Arabic version with the original version will be given to an expert committee (consisting of at least three people). The committee will have experience in the fourth grade for the target Saudi population and excellent knowledge of both languages and the country's cultural context in order to assess the translation's readability and accuracy for that specific population. The committee, comprised of the translators and health and language

professionals, will meet with the purpose of consolidating the different versions of the measurement to produce a final form and ensure equivalence between the source and new versions.

5. **Pretesting:** A sample of 30-40 individuals from the target population will be given the translated material as a pretest by using standard cognitive interviewing techniques.

6. **Submission and Appraisal:** The researchers will document all processes of translation and provide sample forms from different translators, the experts' comments and questions, and the final decisions made by the review committee.

In their article, Forsyth, Kudela, Levin, Lawrence, and Willis (2006) recommended five steps for using the experts' method to translate from language to another:

1. **Translation:** Three translators independently translate the survey from the original language into the target language and document all challenges that they have faced.

    2   **Consultation:** A language consultant will be hired to review the new translations and coordinate pretesting activities.

    3   **Revised:** The translated survey will be given to a team of reviewers to identify problems that they found, document their findings, and suggest revisions.

    4   **Document:** All the processes will be written and documented.

    5.   **Pretested:** The final survey will be pretested using cognitive interviews, and the results will be used to make final changes to the questionnaire.

    In their version of translation and verification, the IEA center used the Expert Translation Method (ETM). In this method, translators were asked to translate a survey or assessment. In their version, a group of translators producing one translation version in a traditional manner, where the translators simply produce a translation to the best of their abilities. The direct approach is the simplest and cheapest translation approach (Harkness and Schoua-

Glusberg,1998). The IEA translated the test and questionnaires to 58 languages (Mullis et al., 2009). Therefore, the center used the 'direct translation', where the items were translated from the source language to the target languages "that is, 'one way' (forward) translation as opposed to 'two way' (forward and backward or 'double') translation, ie., translation and back translation." (Harkness and Schoua-Glusberg,1998,p.101). If the Backward Translation Method (BTM) had been used (at least two independent translators at each step) and the committee and modified committee translation were included, the IEA would have spent more time and money but with high quality. According to Harkness and Schoua-Glusberg(1998) Committee or parallel translation involves several translators who make independent translations of the same questionnaire. At a reconciliation (consensus, revision) meeting, translators and a translation coordinator compare the translations, reconcile discrepancies and agree on a final version which taps the best of the independent translations or, alternatively, appears in the course of discussion." (Harkness and Schoua-Glusberg,1998,p.101). When using the BTM, we can assess the quality of a translation "for survey translation, back translation is seen as offering a solution to the fact that researchers often need information about the quality of translations without being able to read and evaluate these themselves."(Harkness and Schoua-Glusberg,1998,p.101)In fact, the more identical the two versions are, the greater the equivalence between the forward version and backward version is considered to be. In short, BTM is considered to have higher quality more than only direct translation, but the BTM cost more time and money. (Harkness and Schoua-Glusberg,1998)

Summary

Translation of an assessment instrument from one language to another language is not easily or readily achieved. The researcher needs to make sure that the new instrument is semantic, conceptual, and normative equivalence to the source language and can be easily read

by the students. This research involves the Saudi's population that do not speak the original

language that the PIRLS survey was designed for population who speak English, so more

attention and procedures should be undertaken in order to make sure that instrument is valid and

reliable for the Saudi's population. The research investigated the BTM to find out to what

degreethe assessment would be equivalent for the target population.

CHAPTER 3
Methodology

The Progress in International Reading Literacy Study (PIRLS) is an international comparative assessment that measures student achievement in reading. Since 2001, PIRLS has been administered every five years (2001, 2006, 2011 and 2016). PIRLS documents worldwide trends in the reading knowledge of fourth grade students as well as schools and teacher practices related to instruction. Fourth grade students complete a reading assessment and a questionnaire that addresses students' attitudes towards reading, motivation and their reading habits. In addition, questionnaires are given to students' teachers and school principals to gather information about students' school experiences in developing reading literacy. For this project, data for 2011 from Saudi Arabia was used. The total sample was 4,216 participants in 2011 and it was the first time that Saudi Arabia participated in PIRLS. Because the PIRLS cognitive items was not available, and we were prohibited from using it, therefore we turned attention to the affective surveys to begin this language translation verification and accuracy process. Thus, the study focused only on three affective measurements at the students' level. Those three measurements measured students' attitude toward reading (6 items), students' motivation toward reading (6 items), and students' confidence in their reading (7 items). The items that were surveyed are as follows:

First Dimension: Reading Attitude
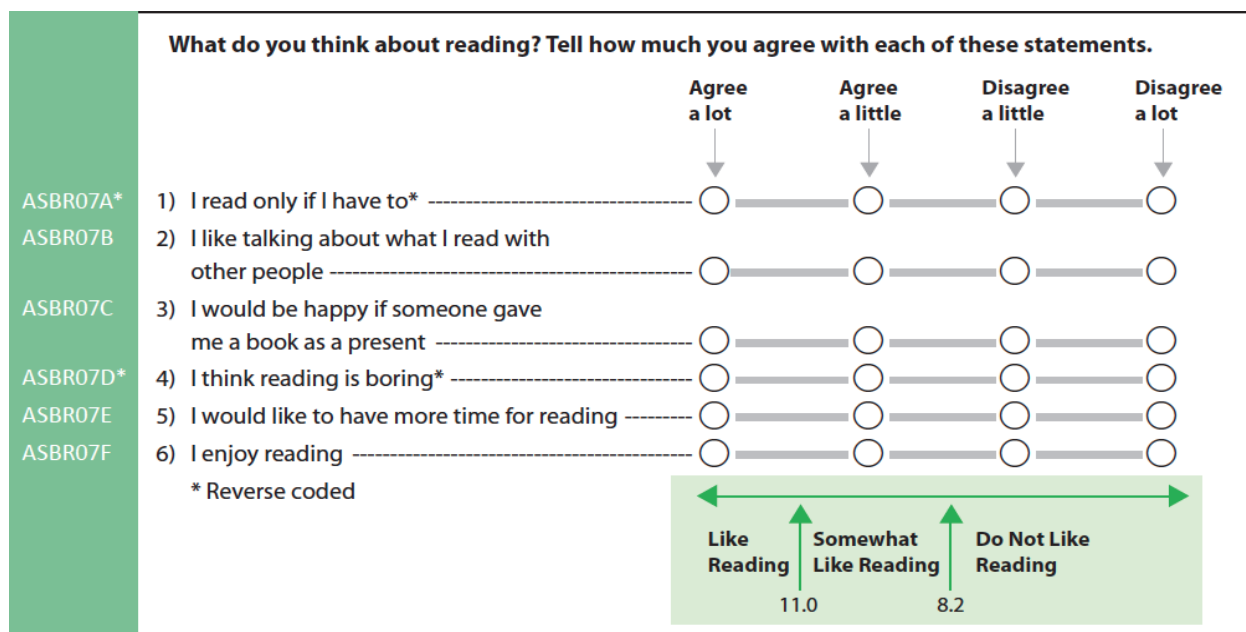
Figure 1: Reading attitude

The reading attitude dimension was measured by 6 items (the range of raw score 0-15). Students who score less than 8.2 were considered to have low attitude toward reading, students who got from 8.2- less than 11 were considered as somewhat like to read and students who got 11 or higher were considered to have high attitude toward reading. The reliability for Students' from Saudi Arabia for this dimension (reading attitude) is .55, which is the index for the items difficulty and discrimination, and it suggests the items are not acceptable for difficulty and discrimination. After calculating the inter correlation between those items, the researcher found that, the mean size of inter-item correlation at .102, which is very weak, and it indicates that the variance for those items is .010.

Second Dimension: Reading Motivation
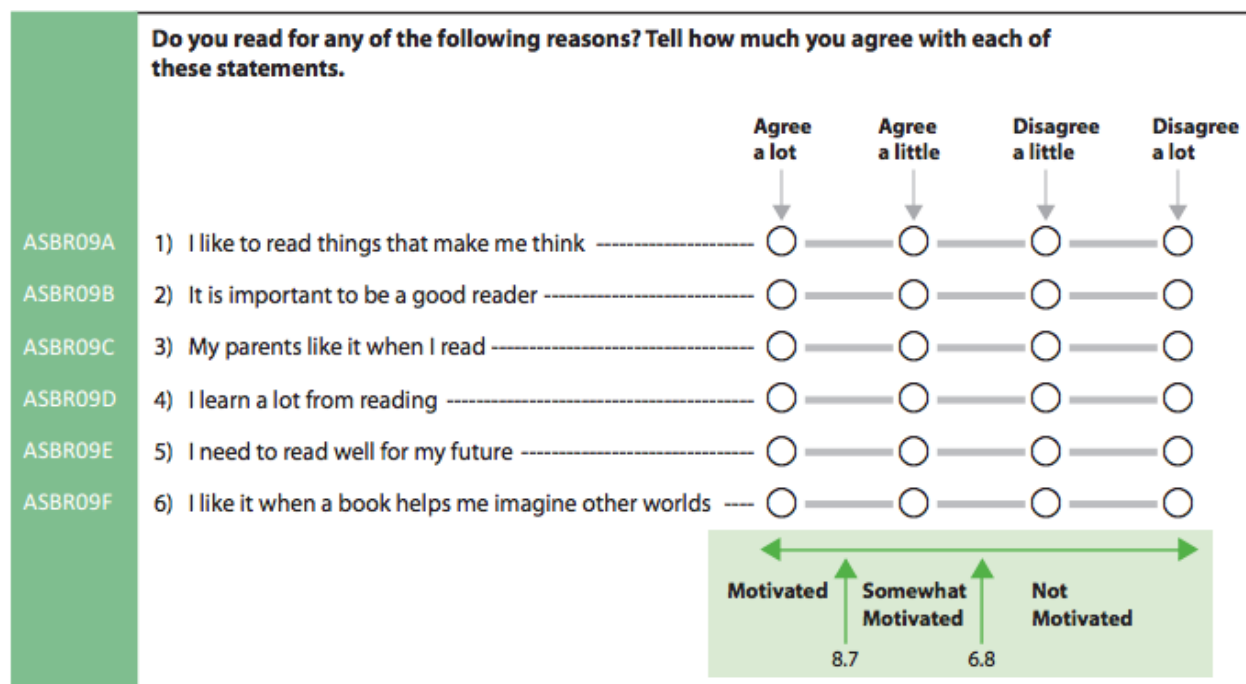
Figure 2: Reading motivation
The reading motivation dimension was measured by 6 items (the range of raw score 0-15).

Students who score less than 6.8 were considered to have low motivation toward reading,

students who got from 66.8- less than 8.7 were considered as somewhat motivated to read and

students who got 8.7 or higher were considered to have high motivation toward reading. The

reliability for Students' from Saudi Arabia for this dimension (reading motivation) is .7, which

is an index for the items difficulty and discrimination, and it suggests the items are not sufficient

for difficulty and discrimination. After calculating the intercorrelation between those items, the

researcher found that, the mean size of inter-item correlation at .306 is weak, and it indicates

that the variance for those items is .128.


Third Dimension: Reading Confidence
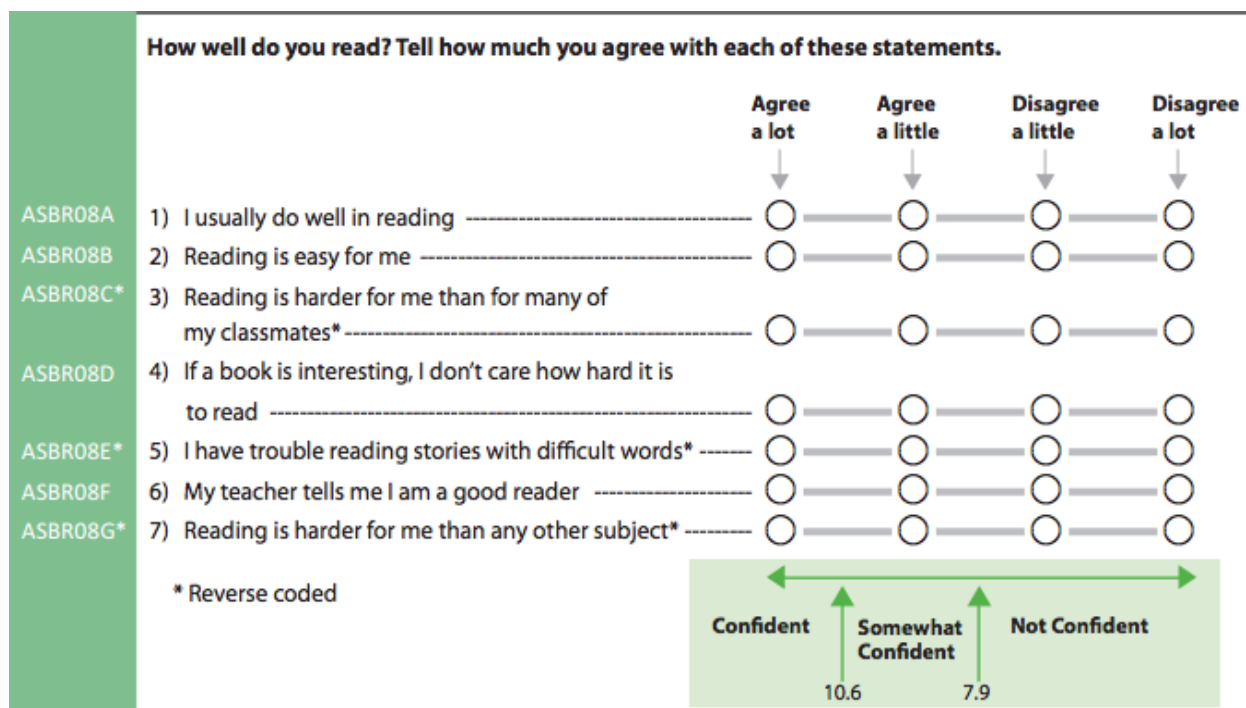
Figure 3: Reading confidence
The reading confidence dimension was measured by 6 items (the range of raw score 0-15).

Students who score less than 7.9 were considered to have low confidence toward reading,

students who got from 7.9- less than 10.6 were considered as somewhat confident to read and

students who got 10.6 or higher were considered to have high confidence toward reading. The

reliability for Students' from Saudi Arabia for this dimension (reading confidence) is .5, which

is an index for the items difficulty and discrimination, and it suggests the items are not

acceptable for difficulty and discrimination. After calculating the inter correlation between those

items, the researcher found that, the mean the size of inter-item correlation at .169 is very weak,

and it indicates that the variance for those items is .002.

Those three dimensions were translated from English to Arabic by team of expert translators and

reviewers , who have the following characteristics (similar to the IEA's experts characteristics) :

- Excellent knowledge of English

- Excellent knowledge of the target language

- Experience in the country's cultural context

- Experience translating texts in the subject areas related to the TIMSS and PIRLS assessments (mathematics, science, and literary texts, respectively)

The final translated Arabic version used for the Saudi sample is in Index A for the three construct. Saudi Arabia used the same items, and they did not add or remove any items among those three constructs. The researcher will use another method of translation, which is Backward Translation Method (BTM), to confirm the PIRLS works, where Experts Translation Methods (ETM) was used, in all these three dimensions.

Procedures

The researcher did BTM for the three constructs following the procedure for each construct as set out by Beaton et al. (2002):

1. **Translation Stage:** At this stage, the researcher invited two independent forward translations, and both of them worked independently. Also, because of the language expression and to control for the gender differences, one of those independent translators was female and the other was male. Both of the translators have a master degree in education and they had at least five years of teaching students at elementary school. Also, they had excellent knowledge in both English and Arabic languages, experience in the Saudi Arabian's cultural context, and experience in translating texts in the area of reading (Appendix A, B, and C). Next, the translated versions were given to three experts who worked together to verify the translation. The three experts compared the original English version to the translated Arabic version to identify discrepancies indicative of ambiguous wording within the original measurement or other problems. Problems were resolved by these verifying translators.

2. **Back Translation:** Two independent translators, who were blind to the original measurement, were invited to translate the "new" Arabic version into an English version (Appendix D, E, and

F). Those translators worked independently. Also, to control for the gender differences in terms of language expression, one of those independent translators was female and another was male. Both of the translators have a master degree in education and they had at least five years of teaching students at elementary school. Also, those documents were given to three experts who worked together to verify the translation. The three experts compared the new Arabic version to the original English version to identify discrepancies indicative of ambiguous wording within the original measurement or other problems.

3. **The Second Back Translation:** Another two independent translators, who were blind to the original measurement and the first backward translation, were invited to translate back the new second English version into the Arabic (second time) (Appendix G, H, and I). Those translators worked independently. Also, to control for the gender differences in term of language expression, one of those independent translators was female and the other male. Both of the translators have a master degree in education and they had at least five years of teaching students at elementary school. Then, those documents were given to the given to three experts who worked together to verify the translation. The three experts compared the Arabic (second time) version to the original English version to identify discrepancies indicative of ambiguous wording within the original measurement or other problems..

4. **Expert Committee Review:** The first Arabic version along with the second Arabic version (Appendix J, K, and L) was given to a three person expert committee who had experience in the fourth grade for the target Saudi population and excellent knowledge of both languages and the country's cultural context in order to assess the translation's readability and accuracy for that specific population. The translators and language professionals met with the purpose of consolidating the different versions of the measurement to produce a final form and ensure equivalence between the first second versions and the original English version. It must be noted:

these experts recommended having two different forms, one for boys and another for girls, because both groups are addressed in different ways in the Arabic language.

The final Arabic versions were prepared for boys (Appendix M, N, and O) and for girls (Appendix P, Q, and R). The committee compared this version to the version that was developed by IEA and answered the following questions:

- Does the BTM differ from the IEA's ETM for Saudi students' reading motivation in the PIRLS assessment?

- Does the BTM differ from the IEA's ETM for Saudi students' reading attitude in the PIRLS assessment?

- Does the BTM differ from the IEA's ETM for Saudi students' reading confidence in the PIRLS assessment?

CHAPTER 4
Results

In the results section, the researcher details the production of the translated English version (English Translated IEA) to the original English version (IEA English), and the Arabic translated version to the IEA Arabic version (IEA translated to Arabic) for each construct. After conducting the BTM, the documents were ready to answer the following questions:

- Does the BTM differ from the IEA's ETM for Saudi students' reading motivation in the PIRLS assessment?

- Does the BTM differ from the IEA's ETM for Saudi students' reading attitude in the PIRLS assessment?

- Does the BTM differ from the IEA's ETM for Saudi students' reading confidence in the PIRLS assessment?

Tables (1-9) show the Cronbach alphas for each of the three Affective scales for Saudi Arabia (English Translated IEA) and United Sates (IEA English) examinees on these scales in 2011 (IEA's data), and for each scale, the item means and item correlations between the scale scores and each item, i.e., item discriminations. Particularly, those tables show how the Cronbach alphas, mean of inter- correlations and correlation between each item to total scores in all three different measurements across the United Sates and Saudi Arabia. The reliability for reading attitude, reading motivation and reading confidence, which are $\alpha= .84$, $\alpha= .77$, and $\alpha= .71$, respectively, for the United States' sample compared to the same nominal Saudi Arabia named scales: $\alpha= .55$, $\alpha= .7$, and $\alpha= .5$, respectively where the items were translated to Arabic language. In fact, the results show that the reading attitude and reading confidence are more reliable for the United States comparing to Saudi Arabia. The processes of translation for those measurements might be a source of this issue (Glas & Jehangir, 2014).

Table 1: Basic Statistics for IEA's data: Saudi Arabia VS. United Sates (Reading Attitude)

Item	Saudi Sample=4163			United States Sample=12083		
	Mean	SD	Correlation to Total	Mean	SD	Correlation to Total
I like to read things that make me think.	2.7	1.243	0.485	2.44	1.202	.587
It is important to be a good reader.	1.72	0.981	0.471	2.29	1.125	.581
My parents like it when I read.	1.38	0.79	0.565	1.98	1.061	.722
I learn a lot from reading.	1.84	1.168	0.641	1.83	1.069	.775
I need to read well for my future.	1.71	0.95	0.59	2.08	1.073	.801
I like it when a book helps	1.47	0.857	0.651	1.70	.986	.840

me to

imagine

other worlds.

	Cronbach's α = .55	Cronbach's α =.84

Table 2:Inter-Item Correlation Matrix (IEA's data For Saudi Arabia, Reading Attitude)

Inter-Item Correlation Matrix (For Saudi Arabia)

	1	2	3	4	5	6
I like to read things that make me think.	1					
It is important to be a good reader.	0.001	1				
My parents like it when I read.	0.051	0.198	1			
I learn a lot from reading.	0.243	0.077	0.233	1		
I need to read well for my future.	0.008	0.183	0.294	0.207	1	
I like it when a book helps me to imagine other worlds.	0.072	0.221	0.361	0.282	0.445	1

Mean Inter-Item Correlations=.102

Table 3:Inter-Item Correlation Matrix (IEA's data For United States, Reading Attitude)

Inter-Item Correlation Matrix (For United States)

	1	2	3	4	5	6
I like to read things that make me think.	1					
It is important to be a good reader.	0.111	1				
My parents like it when I read.	0.222	0.388	1			
I learn a lot from reading.	0.435	0.25	0.445	1		
I need to read well for my future.	0.306	0.369	0.543	0.55	1	
I like it when a book helps me to imagine other worlds.	0.374	0.37	0.538	0.67	0.707	1

Mean Inter-Item Correlations=.418

Table 4: Basic Statistics for IEA's data: Saudi Arabia VS. United Sates (Reading Motivation)

Item	Saudi Sample=4163			United States Sample=12142		
	Mean	Std. Deviation	Correlation to Total	Mean	SD	correlation to Total
I like to read things that make me think.	1.29	.707	.577	1.81	.973	.710
It is important to be a good reader.	1.23	.637	.647	1.27	.625	.685
My parents like it when I read.	1.24	.599	.635	1.33	.677	.623
I learn a lot from reading.	1.27	.633	.695	1.43	.744	.745

I need to read well for my future.	1.22	.606	.677	1.41	.773	.665
I like it when a book helps me imagine other worlds.	1.36	.741	.661	1.46	.838	.686
Cronbach's α = .7				Cronbach's α =.77		

Table 5:Inter-Item Correlation Matrix (IEA's data For Saudi Arabia, Reading Motivation)

Inter-Item Correlation Matrix (For Saudi Arabia)

	1	2	3	4	5	6
I like to read things that make me think.	1					
It is important to be a good reader.	0.216	1				
My parents like it when I read.	0.234	0.292	1			
I learn a lot from reading.	0.262	0.353	0.361	1		
I need to read well for my future.	0.225	0.414	0.347	0.402	1	

I like it when a book	0.24	0.276	0.303	0.354	0.309	1

helps me imagine

other worlds.

Mean Inter-Item Correlations=.306

Table 6:Inter-Item Correlation Matrix (IEA's data For United States, Reading Motivation)

Inter-Item Correlation Matrix (For United States)

	1	2	3	4	5	6
I like to read things that make me think.	1					
It is important to be a good reader.	0.342	1				
My parents like it when I read.	0.293	0.378	1			
I learn a lot from reading.	0.449	0.457	0.375	1		
I need to read well for my future.	0.284	0.436	0.328	0.403	1	
I like it when a book helps me imagine other worlds.	0.384	0.346	0.299	0.406	0.324	1

Mean Inter-Item Correlations=.367

Table 7: Basic Statistics for IEA's data: Saudi Arabia VS. United Sates (Reading Confidence)

Item	Saudi Sample=4163			United States Sample=11972		
	Mean	SD	Correlation to Total	M	SD	Correlation to Total
I usually do well in reading.	1.42	.751	.384	1.50	.727	.655
Reading is easy for me.	1.29	.655	.501	1.49	.761	.707
Reading is harder for me than for many of my classmates.	1.97	1.198	.669	1.76	1.025	.710
If a book is interesting, I don't care how hard it is to read.	1.60	.993	.410	1.62	.935	.428
I have trouble reading stories with difficult words.	2.26	1.209	.613	2.37	1.133	.607

My teacher tells me I am a good reader.	1.49	.861	.466	1.75	.916	.456
Reading is harder for me than any other subject.	1.98	1.226	.655	1.71	1.030	.701
Cronbach's α = .5			Cronbach's α =.71			

Table 8:Inter-Item Correlation Matrix (IEA's data For Saudi Arabia, Reading Confidence)

Inter-Item Correlation Matrix (For Saudi Arabia)

	1	2	3	4	5	6	7
I usually do well in reading.	1						
Reading is easy for me.	0.26	1					
Reading is harder for me than for many of my classmates.	0.073	0.203	1				
If a book is interesting, I don't care how hard it is to read.	0.138	0.235	0.05	1			
I have trouble reading stories with difficult words.	0.06	0.106	0.359	0.036	1		

My teacher tells me I am a

good reader.   0.176   0.29   0.111   0.184   0.1   1

Reading is harder for me

than any other subject.   0.065   0.15   0.453   0.026   0.355   0.112   1

Mean Inter-Item Correlations=.169

Table 9:Inter-Item Correlation Matrix (IEA's data For United States, Reading Confidence)

Inter-Item Correlation Matrix (For United States)

	1	2	3	4	5	6	7
I usually do well in reading.	1						
Reading is easy for me.	0.571	1					
Reading is harder for me than for many of my classmates.	0.336	0.428	1				
If a book is interesting, I don't care how hard it is to read.	0.201	0.256	0.086	1			
I have trouble reading stories with difficult words.	0.221	0.255	0.402	0.091	1		

My teacher tells me I

am a good reader.   0.323   0.274   0.128   0.135   0.055   1

Reading is harder for

me than any other

subject.   0.348   0.4   0.553   0.106   0.365   0.129   1

Mean Inter-Item Correlations=.270

The comparisons done by the committee were based on three main issues:

- Semantic Equivalence: Refers to the words and sentence structure in the translated text

  expressing the same meaning as the source language.

- Conceptual Equivalence: When the concept being measured is the same across groups, although

  wording to describe it may be different.

- Normative Equivalence: Describes the ability of the translated text to address social norms that

  may differ across cultures.

Table 10:*English-Translated Statements (Research Version) vs Original English Statements (English Translated IEA) for Reading Attitude*

	IEA English	**English-Translated Statements**
1.	I read only if I have to.	I read only if I have to.
2.	I like talking about what I read with other people.	I like talking about what I read with other people.
3.	I would be happy if someone gave me a book as a present.	I would be happy if someone gave me a book as a present.
4.	I think reading is boring.	I think reading is boring.
5.	I would like to have more time for reading.	I would like to have more time for reading.
6.	I enjoy reading.	I enjoy reading.

The first question, does the BTM differ than IEA's ETM translation for Saudi students' reading motivation in PIRLS assessment, is addressed in Tables 1 and 2.

Table1 shows that both versions, the IEA English and the translated version, are identical, which means the items have been changed when translated from English to Arabic and then from Arabic to English.

Table 11: Arabic-Translated Statements (Backward Version)vs Arabic Statements (IEA) for Reading Attitude

	Arabic-Translated Statements (Backward Version)	IEA translated to Arabic
1.	أنا أقرأ فقط في حال الحاجة لذلك.	أقرأ عندما أكون مجبراً فقط.
2.	أحب الحديث عما قرأت مع الآخرين.	أحب التحدث عما اقرأه مع أشخاص أخرين.
3.	سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية.	سأفرح إذا اهداني أحد كتابا.
4.	اعتقد بأن القراءة مملة.	أعتقد بأن القراءة أمر ممل.
5.	أود الحصول على مزيداً من الوقت للقراءة.	بودي أن يكون لي مزيد من الوقت للقراة.
6.	أنا استمتع بالقراءة	أنا استمتع بالقراءة

Items 1-5 in Table 2 were translated to the same meaning but in different vocabulary and sentence instructions, so they met the semantic, conceptual, and normative equivalence to the source language. However, for the IEA translated to Arabic, the expert committee who were familiar with the students at the fourth grade found the vocabulary that was used to build those items was above the students' level. The vocabulary was difficult for students to understand because they did not have this level of vocabulary or the sentence instructions in their curricula at fourth grade. However, the new Backward translated version that was carried out in this research had simpler vocabulary and sentence instructions that could be read by students in the middle of the third grade. The vocabulary and the sentence instruction used for this version had been used widely in the students' curricula in the third grade. For this construct, only item 6 was identical for both versions and had semantic, conceptual, and normative equivalence to the

source language and can be easily read by the students in the fourth grade. For this measurement,

about 80% of the items were above the students' level,

Table 12: English-Translated Statements (Research Version) vs Original English Statements (English Translated IEA) for Reading Motivation

	IEA English	**English-Translated Statements**
1.	I like to read things that make me think.	I like to read things that make me think.
2.	It is important to be a good reader.	It is important to be a good reader.
3.	My parents like it when I read.	My parents like it when I read.
4.	I learn a lot from reading.	I learn a lot from reading.
5.	I need to read well for my future.	I need to read well for my future.
6.	I like it when a book helps me imagine other worlds.	I like it when a book helps me imagine other worlds.

Table 3 shows that both versions from the IEA English version and the translated version

are identical, which means the items did not change when translated from English to Arabic and

then from Arabic to English.

Table 13 :*Arabic-Translated Statements (**Backward Version**)vs Arabic-Translated Statements (IEA) for Reading Motivation*

	Arabic-Translated Statements (Backward Version)	IEA translated to Arabic
1.	‎.أحب قراء الأشياء التي تجعلني أفكر	‎.أحب أن اقرأ أمورا تجعلني أفكر
2.	‎.من المهم أن أكون قارئ جيد	‎.من المهم أن أكون قارئا جيدا
3.	‎.والديَ يحبون أن اقرأ بإستمرار	‎.يفرح والدي عندما يجداني أقرأ
4.	‎.أتعلم كثيراً من القراءة	‎.أتعلم الكثير من القرأة
5.	‎أحتاج أن أقرأ جيدا من أجل مستقبلي	‎.علي أن أقرأ جيدا من أجل مستقبلي
6.	‎.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى	‎يعجبني حين يساعدني الكتاب أن اتخيل أماكن أخرى.

Now, consider the translations in Table 4 showing the Backward translated and the IEA

translation to Arabic. For this construct, Reading Motivation, items 1, 2, 4, and 5 were translated

in the same way, and they were found to be semantically, conceptually, and normatively

equivalent to the source language. Also, they can be read by a student in the middle of the third

grade. However, items 3 and 6 in the IEA translation to Arabic are not semantically equivalent to the source language. In addition, these two items do not represent the same meaning in the source language, while in the research version, those two items have the semantic equivalence to the source language as well as the other four items. For this measurement, about 20% of the items had a sematic equivalent issue.

Table 14:*English-Translated Statements (Research Version) vs Original English Statements (English Translated IEA) for Reading Confidence*

	IEA English	**Arabic-Translated Statements**
1.	I usually do well in reading.	I usually do well in reading.
2.	Reading is easy for me.	Reading is easy for me.
3.	Reading is harder for me than for many of my classmates.	Reading is harder for me than for many of my classmates.
4.	If a book is interesting, I don't care how hard it is to read.	If a book is interesting, I don't care how hard it is to read.
5.	I have trouble reading stories with difficult words.	I have trouble reading stories with difficult words.
6.	My teacher tells me I am a good reader.	My teacher tells me I am a good reader.
7.	Reading is harder for me than any other subject.	Reading is harder for me than any other subject.

Table 5 shows that that both versions from the IEA English version and the translated version are identical, which means the items did not change when translated from English to Arabic and then from Arabic to English.

Table 15:*Arabic-Translated Statements (Backward Version1) vs Arabic-Translated Statements (IEA) for Reading Confidence*

	Arabic-Translated Statements (Backward Version)	IEA translated to Arabic
1.	.عادة أجيد القراءة	.عادة أجيد القراءة
2.	.القراءة سهلة بالنسبة لي	.القراءة امر سهل بالنسبة لي
3.	.القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي	القراءة أصعب بالنسبة لي مما هي للكثير من زملائي في الصف.
4.	لا يهمني صعوبة القراءة ما دام الكتاب مشوق	إذا كان الكتاب مشوقا، لا يهمني إن كان صعبا للقراءة.
5.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.	أواجه صعوبة في قراءة قصص تحتوي على كلمات صعبة.
6.	.معلمي يخبرني بأنني قارئ جيد	.تقول معلمتي إنني قاريء جيد

7.	.القراءة أصعب بالنسبة لي من بقية المواد الأخرى	.القراءة أصعب بالنسبة لي من كل موضوع آخر

Table 6 presents the two Arabic translations (Backward and IEA versions) for the Reading Confidence scale. For this item, items 3, 4, and 5 were translated with the same meaning but with different vocabulary and sentence instructions, so they met the semantic, conceptual, and normative equivalence to the source language. However, for the IEA version, the expert committee who were familiar with students in the fourth grade believed the vocabulary, which was used to build these items, was above the students' reading skills level. In addition, the vocabulary was difficult for those students to understand because they did not have this level of vocabulary and the sentence instructions in their curricula in fourth grade. However, the Backward version that was created through this research had simpler vocabulary and sentence instructions that could be comprehended by the students in the middle of the third grade. The vocabulary and the sentence instruction used for this version had been used widely in the students' curricula at the third grade level. Item 7 in the IEA translation to Arabic was not semantically equivalent to the source language. In addition, Item 7 does not represent the same meaning in the source language, while in the research version, the items had semantic equivalence to the source language. Continuing, items 1 and 2 were translated in the same way, and they are semantically conceptually, and normatively equivalent to the source language. Also, they could be read by a student in the middle of the third grade. For this measurement, about 40% of the items had translation issues in IEA translation to Arabic (28% related to reading difficulty and 12% had a semantic equivalent issues).

In summary, results showed that about 50% of the IEA translation to Arabic items had translation issues. In addition, the results show that 40% of the items were considered to be too difficult for students to understand because they were above the students' grade level. Further and very

meaningfully, about 10 % of the items did not represent the same meaning of the source language. The vocabulary that was used for five (5) of the six (6) items on the reading motivation, and three (3) of the seven (7) items on the reading confidence scale were difficult for students to understand because it was above the students' fourth grade skill level. Also, two (2) items in the reading motivation, one (1) item on the reading confidence did not have the same meaning when compared between the translated language and the source language.

CHAPTER 5

Discussion and Conclusion

This study focused on the quality and accuracy of the translations carried out by the IEA (ETM) on the fourth grade assessments of the Progress in International Reading Literacy Study (PIRLS) assessments for Saudi Arabian students in 2011. This study only focused on the quality of translation for three affective constructs from the PIRLS students' attitude survey (reading motivation, reading attitude, and reading confidence). This study used BTM (Beaton et al.2002), from English to Arabic and then from English to Arabic, as the method for translation validation to evaluate whether having the same final Arabic version for those three constructs in three main areas could affect the validity of meaning for the translations. Those three construct areas are semantic equivalence, conceptual equivalence, and normative equivalence.

After comparing between the IEA's ETM and this study's BTM, the study found that about 50% of the IEA translation to Arabic items had translation issues. In addition, about 40% of the IEA translation items were considered to be too difficult for students to understand because they were above the students' grade level. Further and very meaningfully, about 10 % of the items did not represent the same meaning of the source language.

Moreover, the vocabulary that was used by IEA for five (5) of the six (6) items on the reading motivation, and three (3) of the seven (7) items on the reading confidence scale were difficult for students to understand because it was above the students' fourth grade skill level. The lexical difficulty for the Arabic version of those items was higher than the English version, which led to the difficulty of understanding the items when students read the Arabic version. One source of the high lexical difficulty might be that the translators did not use the appropriate Arabic phrasal verbs into English. According to Grisay, Gonzales, and Monseur (2009), "Major cross-language differences in factors related to reading can make it difficult to maintain

cognitive requirement equivalence of the test items, thus affecting their relative difficulty for students assessed in different languages"(p.63). In fact, the item difficulty differential functioning has not been tested to evaluate if all items have similar difficulty across the two comparison countries (e.g., Saudi Arabia, the focal country, and United States, the target country).

Another issue that was discovered while comparing between the two Arabic versions (BTM and ETM versions) is that certain items that translated from Arabic to English by IEA were not semantically equivalent. For example, items for the reading motivation and items for reading confidence did not have the same meaning when compared between the translated language and the source language. This indicates a semantic equivalence issue, where the words and the structure of those items in the translated text do not have the same meaning and expression as the source language. In fact, when translating a construct from English to Arabic language, it is important to be certain that the meaning of each item is the same in both languages, otherwise, that specific item cannot be considered to measure the same construct.(Alonso et al., 1998)

One signal to consider in particular is the reliability for reading attitude, reading motivation and reading confidence, which are α= .84, α= .76, and α= .71, respectively, for the United States' sample compared to the same nominal Saudi Arabia named scales: α= .55, α= .7, and α= .5, respectively where the items were translated to Arabic language. According to Thorndike (1973), there is a "slight tendency for the reliabilities to be higher in the English-speaking countries" (1973, p. 54). The rather great separation observed in this study means that it is necessary to have a new translated version that will fix issues.

Implications

International educational assessments are administrated in many countries, including Saudi Arabia, to inform decisions regarding educational policy, instructional and curricular programs, curriculum evaluation, and new program development. Measurements are used to compare the educated students from one country to other countries, so it is critically important to make sure assessments are equivalent across countries and cultures. Therefore, one of the most important results from this research is that policy makers and   researchers need to be extremely cautious and assured that when using the results of the three assessments (reading attitude, reading motivation, and reading confidence) all items are appropriate for the student population not just in in Saudi Arabia, but all countries. Therefore, as a new revision (BTM) develops and emerges assessment items and questions must be critically reviewed by representative (by age, gender, status, etc.) language and on-grade education instruction experts using multiple translation methods with the intent of making the assessments appropriate and suitable, i.e., content and construct valid, for each participating specific country population. In fact, in their manual book, IEA gives all countries the freedom to add or develop items. Therefore, the Ministry of Education in Saudi Arabia and all other countries are strongly advised for future administrations of the Progress in International Reading Literacy Study (PIRLS) assessments, and all other international normative (relative), growth and criterion (absolute) -referenced tests to embrace the specific recommendations derived from this investigation. The proper and accurate conclusions about the adequacy of intra- and inter- country Education rest in the balance.

Conclusion

The PIRLS assessment instruments were developed by IEA are originally readied in English, and which are then translated into participating countries languages of instruction. IEA seeks a high quality of translations, so they use multiple rounds reviews by linguistic and

assessment experts (ETM) to make sure that the original version, again, which is written in English, is equivalent to the national language(s) of each participating country (Mullis et al., 2009). However, the reliability for the three measurements (reading attitude, reading motivation and reading confidence) studied were very low. After using the BTM, the researcher found that about 50% the IEA translation of the items had translation issues. Therefore, the education practitioners, researchers, policy makers and constituents should consider those issues if they want to use the Saudi Arabia's data for the reading attitude, reading motivation and reading confidence, especially, if the researchers want to compare across countries and cultures. Comparability is required and essential to fairness to create a scale to measure the same constructs across groups being compared with equivalent and high levels of certainty. It has been shown that the use and the interpretation of the assessment scores are the main issue for validity and measurements need to provide evidence to assert their validity across different countries.

Limitations

Because of time, cost and permission realities, this study did not have an opportunity to collect data or interview students to improve and strengthen the validity and the reliability of the translated items and thus scores. Also, the researcher did not have access to the actual test items for Arabic version to include it this study to verify the translations.

References

Alonso, J., Black, C., Norregaard, J.-C., Dunn, E., Andersen, T. F., Espallargues, M., . . . Anderson, G. F. (1998). Cross-cultural differences in the reporting of global functional capacity: an example in cataract patients. *Medical Care, 36*(6), 868-878.

Arnold, B. R., & Matus, Y. E. (2000). Test translation and cultural equivalence methodologies for use with diverse populations. *Handbook of multicultural mental health: Assessment and treatment of diverse populations*, 121-136.

Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2002). Recommendations for the cross-cultural adaptation of health status measures. *New York: American Academy of Orthopaedic Surgeons*, 1-9.

Church, A. T. (2010). Measurement issues in cross-cultural research. *The Sage handbook of measurement*, 151-177.

Cook, L. (2006). *Practical considerations in linking scores on adapted tests.* Paper presented at the Keynote address at the 5th international meeting of the International Test Commission, Brussels, Belgium.

Elley, W. B. (2005). How TIMSS-R contributed to education in eighteen developing countries. *Prospects, 35*(2), 199-212.

Forsyth, B. H., Kudela, M. S., Levin, K., Lawrence, D., & Willis, G. B. (2006). Improving questionnaire translations and translation processes. *Proceedings of Q*.

Glas, C., & Jehangir, K. (2014). Modeling country-specific differential item functioning. *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 97-115.

Greenfield, P. M., & Suzuki, L. K. (2006). Cultural conceptions of learning and development. *Handbook of educational psychology II*, 675-692.

Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. *IERI monograph series: Issues and methodologies in large-scale assessments, 2*, 63-84.

Hambleton, R. K. (2001). The next generation of the ITC Test Translation and Adaptation Guidelines. *European Journal of Psychological Assessment, 17*(3), 164.

Harkness, J., Pennell, B. E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. *Methods for testing and evaluating survey questionnaires*, 453-473.

Harkness, J. A., & Schoua-Glusberg, A. (1998). Questionnaires in translation. ZUMA-Nachrichten Spezial, 7, 87–125.

Harzing, A.-W. (2005). Does the use of English-language questionnaires in cross-national research obscure national differences? *International Journal of Cross Cultural Management, 5*(2), 213-224.

Hurtado, M. P., Angeles, J., Blahut, S. A., & Hays, R. D. (2005). Assessment of the equivalence of the Spanish and English versions of the CAHPS® hospital survey on the quality of inpatient care. *Health services research, 40*(6p2), 2140-2161.

Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review, 54*(1), 5-25.

Kim, M., Han, H.-R., & Phillips, L. (2003). Metric equivalence assessment in cross-cultural research: Using an example of the Center for Epidemiological Studies–Depression Scale. *Journal of Nursing Measurement, 11*(1), 5-18.

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies, 26*(3), 573-596.

Mullis, I. V., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 Assessment framework*: ERIC.

Pan, Y., & De La Puente, M. (2005). Census Bureau guideline for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed. *Survey Methodology, 6.*

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*(4), 1255-1264.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667-706.

Rogers, W. T., Gierl, M. J., Tardif, C., Lin, J., & Rinaldi, C. (2003). Differential validity and utility of successive and simultaneous approaches to the development of equivalent achievement tests in French and English. *Alberta Journal of Educational Research, 49*(3), 290-304.

Rutkowski, D., Rutkowski, L., & von Davier, M. (2013). A Brief Introduction to Modern International Large-Scale Assessment. *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*, 1.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142-151.

Sandilands, D., Oliveri, M. E., Zumbo, B. D., & Ercikan, K. (2013). Investigating sources of differential item functioning in international large-scale assessments using a confirmatory approach. *International Journal of Testing, 13*(2), 152-174.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies, 26*(3), 597-619.

Thorndike, R. L. (1973). Reading Comprehension Education in Fifteen Countries: An Empirical Study. International Studies in Evaluation, III.

Wagemaker, H. (2013). International large-scale assessments: From research to policy. *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*, 11.

APPENDICES

Appendix A
From English to Arabic for Reading Attitude

	Original Statements	Arabic-Translated Statements
1.	I read only if I have to.	.أنا أقرأ فقط في حال الحاجة لذلك
2.	I like talking about what I read with other people.	.أحب الحديث عما قرأت مع الآخرين
3.	I would be happy if someone gave me a book as a present.	.سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية
4.	I think reading is boring.	.اعتقد بأن القراءة مملة
5.	I would like to have more time for reading.	.أود الحصول على مزيداً من الوقت للقراءة
6.	I enjoy reading.	أنا استمتع بالقراءة

Appendix B
From English to Arabic for Reading Motivation

	Original Statements	Arabic-Translated Statements
1.	I like to read things that make me think.	.أحب قراءة الأشياء التي تجعلني أفكر
2.	It is important to be a good reader.	.من المهم أن أكون قارئ جيد
3.	My parents like it when I read.	.والديَ يحبون أن اقرأ بإستمرار
4.	I learn a lot from reading.	.أتعلم كثيراً من القراءة
5.	I need to read well for my future.	أحتاج أن أقرأ جيدا من أجل مستقبلي
6.	I like it when a book helps me imagine other worlds.	.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى

Appendix C
From English to Arabic for Reading Confidence

	Original Statements	Arabic-Translated Statements
1.	I usually do well in reading.	عادة أجيد القراءة.
2.	Reading is easy for me.	القراءة سهلة بالنسبة لي.
3.	Reading is harder for me than for many of my classmates.	القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.
4.	If a book is interesting, I don't care how hard it is to read.	لا يهمني صعوبة القراءة ما دام الكتاب مشوق
5.	I have trouble reading stories with difficult words.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.
6.	My teacher tells me I am a good reader.	معلمي يخبرني بأنني قارئ جيد.
7.	Reading is harder for me than any other subject.	القراءة أصعب بالنسبة من بقية المواد الأخرى.

Appendix D
From Arabic to English for Reading Attitude

	Arabic-Translated Statements	English-Translated Statements
1.	.أنا أقرأ فقط في حال الحاجة لذلك	I read only if I have to.
2.	.أحب الحديث عما قرأت مع الآخرين	I like talking about what I read with other people.
3.	.سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية	I would be happy if someone gave me a book as a present.
4.	.اعتقد بأن القراءة مملة	I think reading is boring.
5.	.أود الحصول على مزيداً من الوقت للقراءة	I would like to have more time for reading.
6.	أنا استمتع بالقراءة	I enjoy reading.

Appendix E
From Arabic to English for Reading Motivation

	Arabic-Translated Statements	English-Translated Statements
1.	.أحب قراءة الأشياء التي تجعلني أفكر	I like to read things that make me think.
2.	.من المهم أن أكون قارئ جيد	It is important to be a good reader.
3.	.والديَ يحبون أن اقرأ بإستمرار	My parents like it when I read.
4.	.أتعلم كثيراَ من القراءة	I learn a lot from reading.
5.	أحتاج أن أقرأ جيدا من أجل مستقبلي	I need to read well for my future.
6.	.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى	I like it when a book helps me imagine other worlds.

Appendix F
From Arabic to English for Reading Confidence

	Arabic-Translated Statements	English-Translated Statements
1.	.عادة أجيد القراءة	I usually do well in reading.
2.	.القراءة سهلة بالنسبة لي	Reading is easy for me.
3.	.القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي	Reading is harder for me than for many of my classmates.
4.	لا يهمني صعوبة القراءة ما دام الكتاب مشوق	If a book is interesting, I don't care how hard it is to read.
5.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.	I have trouble reading stories with difficult words.
6.	.معلمي يخبرني بأنني قارئ جيد	My teacher tells me I am a good reader.
7.	.القراءة أصعب بالنسبة من بقية المواد الأخرى	Reading is harder for me than any other subject.

Appendix G
From translated English to Arabic for Reading Attitude

	English-Translated Statements	Arabic-Translated Statements (version2)
1.	I read only if I have to.	.أنا أقرأ فقط في حال الحاجة لذلك
2.	I like talking about what I read with other people.	.أحب الحديث عما قرأت مع الآخرين
3.	I would be happy if someone gave me a book as a present.	.سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية
4.	I think reading is boring.	.اعتقد بأن القراءة مملة
5.	I would like to have more time for reading.	.أود الحصول على مزيداً من الوقت للقراءة
6.	I enjoy reading.	أنا استمتع بالقراءة

Appendix H
From translated English to Arabic for Reading Motivation

	English-Translated Statements	Arabic-Translated Statements (version2)
1.	I like to read things that make me think.	.أحب قراءة الأشياء التي تجعلني أفكر
2.	It is important to be a good reader.	.من المهم أن أكون قارئ جيد
3.	My parents like it when I read.	.والديَ يحبون أن اقرأ بإستمرار
4.	I learn a lot from reading.	.أتعلم كثيراً من القراءة
5.	I need to read well for my future.	أحتاج أن أقرأ جيدا من أجل مستقبلي
6.	I like it when a book helps me imagine other worlds.	.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى

Appendix I
From translated English to Arabic for Reading Confidence

	English-Translated Statements	Arabic-Translated Statements (version2)
1.	I usually do well in reading.	عادة أجيد القراءة.
2.	Reading is easy for me.	القراءة سهلة بالنسبة لي.
3.	Reading is harder for me than for many of my classmates.	القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.
4.	If a book is interesting, I don't care how hard it is to read.	لا يهمني صعوبة القراءة ما دام الكتاب مشوق
5.	I have trouble reading stories with difficult words.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.
6.	My teacher tells me I am a good reader.	معلمي يخبرني بأنني قارئ جيد.
7.	Reading is harder for me than any other subject.	القراءة أصعب بالنسبة من بقية المواد الأخرى.

Appendix J
Arabic-Translated Statements (version1) VS. Arabic-Translated Statements (version2) for Reading Attitude

	Arabic-Translated Statements (version1)	Arabic-Translated Statements (version2)
1.	أنا أقرأ فقط في حال الحاجة لذلك.	أنا أقرأ فقط في حال الحاجة لذلك.
2.	أحب الحديث عما قرأت مع الآخرين.	أحب الحديث عما قرأت مع الآخرين.
3.	سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية.	سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية.
4.	اعتقد بأن القراءة مملة.	اعتقد بأن القراءة مملة.
5.	أود الحصول على مزيداً من الوقت للقراءة.	أود الحصول على مزيداً من الوقت للقراءة.
6.	أنا استمتع بالقراءة	أنا استمتع بالقراءة

Appendix K
Arabic-Translated Statements (version1) VS. Arabic-Translated Statements (version2) for
Reading Motivation

	Arabic-Translated Statements (version1)	Arabic-Translated Statements (version2)
1.	.أحب قراءة الأشياء التي تجعلني أفكر	.أحب قراءة الأشياء التي تجعلني أفكر
2.	.من المهم أن أكون قارئ جيد	.من المهم أن أكون قارئ جيد
3.	.والديَ يحبون أن اقرأ بإستمرار	.والديَ يحبون أن اقرأ بإستمرار
4.	.أتعلم كثيراً من القراءة	.أتعلم كثيراً من القراءة
5.	أحتاج أن أقرأ جيدا من أجل مستقبلي	أحتاج أن أقرأ جيدا من أجل مستقبلي
6.	.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى	.أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى

Appendix L
Arabic-Translated Statements (version1) VS. Arabic-Translated Statements (version2) for Reading Confidence

	Arabic-Translated Statements (version1)	Arabic-Translated Statements (version2)
1.	عادة أجيد القراءة.	عادة أجيد القراءة.
2.	القراءة سهلة بالنسبة لي.	القراءة سهلة بالنسبة لي.
3.	القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.	القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.
4.	لا يهمني صعوبة القراءة ما دام الكتاب مشوق	لا يهمني صعوبة القراءة ما دام الكتاب مشوق
5.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.	أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.
6.	معلمي يخبرني بأنني قارئ جيد.	معلمي يخبرني بأنني قارئ جيد.
7.	القراءة أصعب بالنسبة من بقية المواد الأخرى.	القراءة أصعب بالنسبة من بقية المواد الأخرى.

Appendix M
Final Arabic-Translated Statements for Reading Attitude (For Boys)

موافق بشدة	موافق	معارض	معارض بشدة		
				أنا أقرأ فقط في حال الحاجة لذلك.	1.
				أحب الحديث عما قرأت مع الآخرين.	2.
				سأكون سعيدًا إذا قدم لي أحدهم كتابًا كهدية.	3.
				اعتقد بأن القراءة مملة.	4.
				أود الحصول على مزيداً من الوقت للقراءة.	5.
				أنا استمتع بالقراءة	6.

Appendix N
Final Arabic-Translated Statements for Reading Motivation (For Boys)

موافق بشدة	موافق	معارض	معارض بشدة		
				أحب قراء الأشياء التي تجعلني أَفكر.	1.
				من المهم أن أكون قارئ جيد.	2.
				والديَ يحبون أن اقرأ بإستمرار.	3.
				أتعلم كثيراً من القراءة.	4.
				أحتاج أن أقرأ جيدا من أجل مستقبلي	5.
				أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى.	6.

Appendix O
Final Arabic-Translated Statements for Reading Confidence (For Boys)

موافق بشدة	موافق	معارض	معارض بشدة		
				.عادة أجيد القراءة	1.
				.القراءة سهلة بالنسبة لي	2.
				القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.	3.
				لا يهمني صعوبة القراءة ما دام الكتاب مشوق	4.
				أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.	5.
				.معلمي يخبرني بأنني قارئ جيد	6.
				القراءة أصعب بالنسبة لي من بقية المواد الأخرى.	7.

Appendix P
Final Arabic-Translated Statements for Reading Attitude (For Girls)

موافق بشدة	موافق	معارض	معارض بشدة		
				أنا أقرأ فقط في حال الحاجة لذلك.	1.
				أحب الحديث عما قرأت مع الآخرين.	2.
				سأكون سعيدة إذا قدم لي أحدهم كتابًا كهدية.	3.
				اعتقد بأن القراءة مملة.	4.
				أود الحصول على مزيداً من الوقت للقراءة.	5.
				أنا استمتع بالقراءة	6.

Appendix Q
Final Arabic-Translated Statements for Reading Motivation (For Girls)

موافق بشدة	موافق	معارض	معارض بشدة		
				أحب قراءة الأشياء التي تجعلني أُفكر.	1.
				من المهم أن أكون قارئة جيدة.	2.
				والديَ يحبون أن اقرأ بإستمرار.	3.
				أتعلم كثيراً من القراءة.	4.
				أحتاج أن أقرأ جيدا من أجل مستقبلي	5.
				أستمتع عندما يساعدني الكتاب على تخيل عوالم أخرى.	6.

Appendix R
Final Arabic-Translated Statements for Reading Confidence (For Girls)

موافق بشدة	موافق	معارض	معارض بشدة		
				عادة أجيد القراءة.	1.
				القراءة سهلة بالنسبة لي.	2.
				القراءة أصعب بالنسبة لي مقارنة بكثير من زملائي.	3.
				لا يهمني صعوبة القراءة ما دام الكتاب مشوق	4.
				أجد صعوبة في قراءة القصص التي تحتوي على كلمات صعبة.	5.
				معلمتي تخبرني بأنني قارئة جيد.	6.
				القراءة أصعب بالنسبة لي من بقية المواد الأخرى.	7.