

Adaptation, Translation, and Validation of Information Literacy Assessment Instrument

Nordin Hani Syazillah

Department of Library & Information Science, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia. E-mail: syazillah@siswa.um.edu.my

K. Kiran

Department of Library & Information Science, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia. E-mail: kiran@um.edu.my

G. Chowdhury

Department of Mathematics and Information Sciences, Northumbria University, Newcastle, UK. E-mail: gobinda.chowdhury@northumbria.ac.uk

The assessment of information literacy (IL) at the school level is mainly dependent on the measurement tools developed by the Western world. These tools need to be efficiently adapted and in most cases translated to allow them to be utilized in other cultures, languages, and countries. To date, there have been no standard guidelines to adapt these tools; hence, the results may be cross-culturally generalized to a certain extent. Furthermore, most data analyses produce generic outcomes without taking into account the ability of the students, including the difficulty of the test items. The present study proposes a systematic approach for context adaptation and language translation of the preexisting IL assessment tool known as TRAILS-9 to be used in different languages and context, particularly a Malaysian public secondary school. This study further administers a less common psychometric approach, the Rasch analysis, to validate the adapted instrument. This technique produces a hierarchy of item difficulty within the assessment domain that enables the ability level of the students to be differentiated based on item difficulty. The recommended scale adaptation guidelines are able to reduce the misinterpretation of scores from instruments in multiple languages as well as contribute to parallel development of IL assessment among secondary school students from different populations.

Introduction

A large body of literature has been devoted to the establishment of information literacy (IL) competency standards for different target groups and to the development of related IL assessment tools. However, most measures developed are in the English language and intended for English-speaking countries, despite a few exceptions developed by Rattana-wongsa and Koraneekij (2014) and Korobili, Malliari, and Christodoulo (2009). There are two available options that can be used to assess the IL skills of a particular group: (i) develop a new measure, or (ii) use a previously developed measure (possibly in another language and cultural context). Furthermore, the development of a new test is a complex and difficult process (Gall, Gall, & Borg, 2003, p. 216). As such, the use of a predeveloped instrument is believed to be an advantage to the researcher. The use of an existing instrument not only saves time and energy, but also allows the study to be connected to the entire body of knowledge based on a particular construct (Korb, 2012). However, Dorner and Gorman (2006) cautioned that measures or tests for IL that are performed based on Western norms and taught according to Western pedagogical practices may not be suitable for non-English-speaking countries considering the effects of cultural factors towards the curriculum and program delivery of IL education. On top of that, careful consideration has to be made on the linguistics aspects as well as content validity of the instrument across different cultures (Guillemin, Bombardier, & Beaton, 1993).

According to Walsh (2009), there are nine different types of IL assessment: bibliographies, essays, portfolios, MCQ,

final grade, observations, quiz, self-assessment, and simulation. The multiple-choice question (MCQ) type assessment is reported to be the most popular, considering that it is convenient, quick, and easy to use, especially to test specific knowledge and skills using test items that can be carefully mapped to a particular IL standard. Several multiple-choice tests such as the Information Literacy Test (ILT) by James Madison University and TRAILS (Tool for Real-time Assessment of Information Literacy Skills) by Kent State University Libraries, were carefully developed through vigorous empirical methods and these tests have been repeatedly used by others (Gross & Latham, 2007; Cameron, Wise, & Lottridge, 2007; Chu, 2012). Although the use of adapted instruments contributes to the consistency of an instrument and its measures, the concern is when researchers and practitioners use available measures with little attention to the validity of the measure in different sociocultural domains. Mere translation of items, instructions, and response options does not ensure the retention of validity and reliability of the scale (Beaton et al., 2000). Furthermore, the issue of cross-cultural adaptation of information literacy measures has received very little attention in library and information science (LIS) research. In relation to this, there is no general agreement on how to adapt an IL assessment instrument to a cultural setting that is different from that in which it was developed.

Therefore, the aim of this article is to illustrate the steps involved in a cross-cultural adaptation of a research instrument using TRAILS. This study further utilizes a data analysis tool known as Rasch Modeling, which is not commonly used in library and information science research for the purpose of strengthening the methodology applied in LIS research. The proposed method may serve as a guideline to other researchers when using a preexisting instrument.

Literature Review

The emergence of IL standards and guidelines by AASL, ACRL, CAUL, SCONUL, and CILIP necessitate assessment as a means to measure performance against the standards (Majid, Chang, & Foo, 2016). The notion that states college and university students have higher-order thinking skills and improved abilities has led researchers to experiment with various modes of IL performance assessments such as rubrics (Oakleaf, 2009; van Helvoort, 2013; Gola et al., 2014) and portfolios (Dilller & Phelps, 2008). Several other common methods of assessment include analysis of bibliographies, multiple-choice tests, observations, quiz/tests, self-assessments, and simulation (Walsh, 2009). One of the well-cited works on IL assessment is by Oakleaf (2008), who has contextualized her discussion of IL instructions assessment which range from scientific measures (the use of fixed-choice tests) to constructivist educational theories (the use of qualitative assessment methods). She strongly advocates that librarians understand the theories that underlie the assessment tool they intend to use.

The use of qualitative assessment tools developed by Whitlock and Ebrahimi (2016) is designed for project-based or activity-based IL instructions, which is normally carried out in university libraries for the purpose of addressing higher-level skills. These types of assessment tools are difficult to be adapted to a wider audience because they are highly dependent on the task at hand, such as the project topic, the available resources, and the time constraints. Hence, a quantitative measurement tool is deemed more practical to develop a method of adopting and adapting reliable and valid measuring tools.

A review performed by Walsh (2009) on a total of 91 studies on IL assessment methods revealed that multiple-choice tests are the most commonly used method. In higher education, the most widely used ILT was developed by James Madison University (JMU) (Cameron, Wise, & Lottridge, 2007). Interestingly, the use of ILT by Gross and Latham (2007, 2012, 2013) revealed that there are bottom-tier students who are information proficient as well as top-tier students who are nonproficient. In relation to this, it is interesting to note that students' self-view of their IL skills is always higher than their actual score on the IL test.

In Singapore, Foo et al. (2014) developed a new ILT for Primary 3 students, with the results revealing a low level (below 50/100) of IL skills score. However, the results were not dependable due to the low reliability of Cronbach alpha recorded at 0.5, which indicates only a satisfactory level of internal consistency for the data collection instrument. Hence, further refinement of the test items is deemed necessary.

The lack of readily available IL assessment tools (Kovalik, Yutzey, & Piazza, 2012) has led to the limited number of empirical research being performed on the growth of IL skills throughout students' K-12 education (Latham & Gross, 2008). On top of that, higher-order skills such as information synthesis and information evaluation have become another area of concern because they are often reported as low achieved skills (Foo et al., 2014). Hence, this further leads to the rise of the following questions: (1) Are the items measuring these skills too difficult for the ability level of the students? and (2) Are we assuming that the inability to answer correctly is solely due to the undeveloped students' skills? Therefore, it is important to ensure that the test item accurately operationalizes the domain to be measured as well as matches the ability level of the students.

An assessment does not function in isolation. It is targeted at improving students' learning experiences through the measurement of performance outcomes that are dependent on predetermined indicators or standards. Hence, as suggested by Willer (2014), test items that are newly developed or straightforwardly adapted must represent developmentally appropriate grade level designations.

Several IL assessment tools have been developed by libraries, colleges, and universities to cater to the need of measuring the mastery of IL skills. More important, these established tools are subject to several years of development and continual testing. Several examples of these tests

include ILT, iSkills, SAILS (Standardized Assessment of Information Literacy Skills), and TRAILS. These tools have been repeatedly adopted by researchers with slight modifications and adaptations. In relation to this, a detailed overview of the various IL standards and assessment techniques used in higher education by Boh Podgornik et al. (2015) concludes that most of the existing IL tests are somewhat “library-based” and “dedicated to local use.”

There have also been isolated efforts by non-English-speaking countries to create IL measuring instruments or scales from scratch by referring to the existing standards. A number of countries that created such IL measuring instruments or scales include Turkey (Kurbanoglu, Akkoyunlu, & Umay, 2006), Oman (Al-Aufi & Al-Azri, 2013), Vietnam (Ngo, Walton, & Pickard, 2016), and Singapore (Chang et al., 2012; Foo et al., 2013). However, these new initiatives are often the result of a single study; hence, they are not able to significantly contribute to the development of the knowledge base for IL assessment unless they are widely adopted and retested in varying contexts.

Kurbanoglu et al. (2006) developed an Information Literacy Self-Efficacy Scale (ILSES) for measuring teachers' IL skills. The original instrument was in the Turkish language, but was later translated to English and pretested among bilingual students. The study only managed to detail the empirical testing of the scale without describing how the initial 40 items were derived. Hence, it is difficult for other researchers to fully grasp the conceptual stand and how it contributes to the larger knowledge base.

Translation is an important process in adopting existing research instruments. It is of utmost importance that the validity of the instrument is not compromised in the translation process. In Hong Kong, Chu (2012) translated TRAILS using the Chinese text for Primary 5 students, and it was reported that his study is the first attempt that applies TRAILS for a non-American population.

It is much more viable to adapt an instrument already in use based on the assumption that the process of scale development is time-consuming. However, there is a lack of understanding on how to systematically adapt an instrument developed in a context that is different from the current study. Several studies have adopted the scale as a whole (e.g., Chu, 2012; Baji & Bigdali, 2016; Smith et al., 2013), but there are limitations on the validity of the scale, as these studies only report the reliability scores. A number of well-established protocols for translating, adapting, and validating instruments have been developed (Guillemin, Bombardier, & Beaton, 1993; Beaton et al., 2000; Sousa & Rojjanasrirat, 2011); however, no consensus has been achieved among LIS researchers on how IL tests can be refined to be used in different populations. It is obvious that the LIS literature has not yet provided any guidelines on how existing measurement scales or instruments may be adapted, translated, and validated for widespread use. Hence, from a measurement perspective, it can be concluded that the IL literature in guiding the test adaptation process is incomplete.

Therefore, the purpose of this study is to present a highly recommended methodological approach for translating, adapting, and validating a predeveloped IL assessment instrument in a context other than the originally intended. Similar to other empirical investigation, the IL assessment scope for this study is narrowed to the context of IL among secondary school students.

Research Objectives

The purpose of this study is twofold. First, it addresses the methodological challenges involved in adapting a preexisting instrument into other cultures, languages, and countries. Second, it proposes a less commonly employed psychometric approach known as the Rasch analysis for the purpose of validating IL assessment based on student ability. The objectives of this study are specified as follows:

1. To describe the methodological approach employed for the purpose of standardizing and systematically translating an established IL assessment tool for cross-cultural applicability.
2. To test and validate an adapted and translated IL assessment tool using the Rasch analysis.

Methodological Approach

There are two major phases in this study that lead to the validated IL assessment tool for Malaysian secondary school students. The first phase involves the adaptation of an appropriate tool known as TRAILS to accurately assess the IL level of the selected study sample. The second phase involves the validation of the adapted and translated instrument using Rasch analysis, which is considered a systematic method to categorize individuals according to their ability and items according to their level of difficulty.

In Malaysia, children spend 5 years in the secondary school which is divided into lower secondary (Grades 7–9) and upper secondary (Grades 10–11). In Grade 9, students are required to sit for the national examination which allows them to be assigned to either the Arts or Science stream. Any research involving upper secondary students is limited to Grade 10, as the Ministry of Education does not grant permission to recruit Grade 11 students as respondents. There is no direct inclusion of IL education in the national school curriculum, but efforts to integrate IL through students' assignments are evident, particularly in terms of resource-based projects (Yu et al., 2016) given throughout Grades 7–11. As such, no particular IL standard is referred to when attempting to deliver IL education to secondary school students. Therefore, it is not possible to directly map TRAILS grade level (or any other assessment tool) to Malaysian students, thus making it imperative to adapt any existing measurement using systematic reasoning and procedures.

This study involves two secondary schools that represent the typical conventional schools in Malaysia. As mentioned, the sample is only limited to Grade 10 students. In the first

phase, two sets of 35 students each were selected from School I for item-testing. Class teachers were solicited to pick a mix of high achievers and low achievers to ensure a nonbiased feedback on their understanding and acceptability of the test items. As suggested by Beaton et al. (2000), a total of 30–40 respondents are viewed as appropriate for pre-testing. In the second phase, there were a total of 199 Form Four students available in School II, ranging from high achievers to low achievers. The test instrument was distributed to all 199 respondents with the help of the class teachers; however, only 163 complete responses were returned and utilized for analysis.

The next two sections provide details on each of the phase involved in this study.

The Instrument Adaptation and Translation

TRAILS was chosen for this study considering that it is a multichoice item instrument intended to assess the IL skills based on the American Association of School Librarians' *Standards for the 21st-Century Learner* and those from the Common Core State Standards. The TRAILS project was initiated by Kent State University Libraries in 2004 and its active collaboration with several institutes has seen revisions in the form of TRAILS-9, -6, -12, and the development of TRAILS-3 in 2012 (<http://www.trails-9.org/>). TRAILS measures IL on a continuum of five stages as follows:

1. Develop Topic;
2. Identify Potential Sources;
3. Develop, Use, and Revise Search Strategies;
4. Evaluate Sources and Information; and
5. Recognize how to use Information Responsibly, Ethically, and Legally.

Each assessment consists of a varying number of items by grade. The total score for each student is provided at the end of the test to produce a measure of attainment relative to other students who have taken the same test. Therefore, the score is not a definitive measure of students' IL knowledge. A detailed analysis of the establishment of TRAILS can be found in Schloman and Gedeonoakleaf (2007) and Miller (2016).

The Instrument adaptation and translation involves three major steps as follows:

Identification of test content. First, it is important to define what the test and the test items will measure, including the target domain of knowledge and skills. In this study, the American Association of School Librarians' Standards for the 21st-Century Learner are conceptualized as the necessary IL standard for the Malaysian upper secondary school students. Hence, TRAILS with its five information domains (constructs) is considered a suitable test instrument, considering that it incorporates the operationally defined items for the measure of each construct. TRAILS has four different test instruments that are constructed on student cognitive

ability levels; hence, it is necessary to identify which of the four instruments closely match the context of the learners as well as the learning objectives of Malaysian secondary school students.

A team of five secondary school teachers, each with 5–15 years of teaching experience, were asked to assess TRAILS-3, TRAILS-6, TRAILS-9, and TRAILS-12. They examined the suitability of each grade standards, especially in terms of conceptual, content, and literacy criteria. A consensus among the five teachers resulted in the selection of the *Grade 9 General Assessment 2* (TRAILS-9) as the most accurate tool to assess the performance of Malaysian Grade 10 (aged 16 years) students.

Item translation. The symmetrical translation category is a recommended approach because it retains both meaning and colloquialism in the original language as well as the target language (Sousa & Rojjanasrirat, 2010). Item translation is conducted based on the following steps:

Step 1: Forward translation

Translation of the TRAILS-9 items from English to the Malay language is carried out by two separate entities: a certified bilingual translator and a group of five teachers. The intention is to allow the certified translator to focus on colloquial phrases, slang, and jargon and terms that are commonly used in a Malaysian context. Meanwhile, the teachers focus on the content and knowledge level of the target students. Both translated versions are then brought to the attention of a team for the purpose of addressing any ambiguities and discrepancies of words, sentences, and meaning.

Step 2: Comparison of the translated versions

The team is comprised of: (i) four teachers who are responsible to verify the suitability of the items according to the students' cognitive level; (ii) two university librarians who are assigned to check for information literacy skill elements; and (iii) two LIS professors (experts) who are expected to assess the final version of the instrument for robustness. Several changes were required to be made by the team. Item IP[15] involves a change of direct translated terms: the direct translation of the term up-to-date in Malay is "*terkini*." However, the teachers suggested the use of the phrase [up to date] in brackets considering that it is a phrase that is more familiar to the students. Another comment made by the teachers was on the use of the word "*bibliografi*" [bibliography] in item RH[25]. They suggested using the term "*senarai rujukan*" [reference list] because the students are more familiar with this term in their school project work. In terms of content, a change was performed on item RH[24] in which students have to choose the correct paraphrasing of an excerpt of an article addressing the topic of "drought and winter." The topic in the example was changed to "haze"—a current environment issue in Malaysia. For item RH[25], the teachers suggested that it should refer to the national library reading program (NILAM—*Nadi Ilmu Amalan Membaca*) for the purpose of familiarizing these students with real events

TABLE 1. Adaptation to the TRAILS-9 tool.

Original (TRAILS-9)	Adaptation	Issue
[DT]9: selection of the broadest topic from a list of topics relating to disease	The example given is on the topic of “sports” to illustrate the question	Comprehension level
(DT)11: identifying sources of information. Topic: “fuel cells”	Topic is changed to “Chinese influence on Malaysian architecture” and relevant answers are changed too	Context
[IP]15: .locate <i>up-to-date</i> facts. . .	The direct translation of the term in Malay is “ <i>terkini</i> .” However the teachers suggested use of the word (up to date) in brackets as a word more familiar to students	Colloquial language
[RH]24: which addressed an article relating to “ <i>drought and winter</i> ”	Article was changes to “haze”—a current environmental problem in Malaysia	Context
[RH]25: the title page of a book “Three nights in August”	Title page is change to a local book “ <i>Tiada Jalan Keluar</i> ”	Context
[RH]25: “. . .you are creating a <i>bibliography</i> for. . .	“bibliography” is translated as “ <i>bibliografi</i> ” in Malay. However the teachers suggested that the students are more familiar with the term “ <i>rujukan</i> ” [<i>reference</i>]	Library jargon

conducted at the school. The librarians managed to confirm that the category RH[25] is within the domain of IL. Next, the experts suggested that the title of the book included in item RH[25] should be changed to a title of a local Malay book that is more frequently read by teenagers in Malaysia. Table 1 depicts some of the major changes that were made to the adapted version of TRAILS. This process generated the preliminary translated version of TRAILS-9 in the Malay language.

Step 3: Back translation

The translation from the Malay instrument back to the source language (English) was performed by a second independent certified translator. It is preferable that the translator’s mother tongue is the source language (English), but it was not possible for this study. However, the translator chosen for this study is a qualified Malay–English translator. He accepted the translation without any change.

Item testing. Item testing in this study was performed twice, on a total of 35 secondary school students in each test. First, the Malay version of the instrument was administered to a total of 35 students with the aim of receiving feedback through a dialog session with the researchers. The main concern here was content relevancy and student literary level. Overall, the students did not point out any major difficulty in comprehending the language and content. However, item DT[9] was significantly difficult for the students, as it listed the terms relating to a disease that required the students to identify the broadest term of the topic. After a discussion with the teachers, an example was added to the question in order to assist the students in comprehending the context of the item. The librarians further assured that the example (topic: sports) will not influence the ability of the students to identify the given topic (diseases). Table 1 presents some of the necessary adaptations to the items.

It is important to change the context of the items without jeopardizing the skill measured by the test item. An unfamiliar context may influence the ability of the students to

answer correctly, which turns it into a difficult question. In some instances, library jargon had to be replaced with natural language. Moreover, it is very important that the use of the language does not hinder the students’ understanding of the test item, and in no way impacts their ability to answer the test item correctly.

These steps resulted in a 25-item instrument with the composition depicted in Table 2. A pilot study was then conducted on the second group consisting of 35 students. No issues were raised by the students in answering all 25 items.

Instrument Validation

Test administration. The permission to conduct the survey was obtained from the Ministry of Education, Malaysia, with the limitation that the survey can only be performed on the test group of Form 4 students (Grade 10). The total population of the Form 4 students is 199, in which the final paper-and-pencil adapted TRAILS-9 test was administered during their school term. The students are in the range of high achievers in the first class to the low achievers in the last class of the form. They were briefed on the purpose of the test and informed about the confidentiality of their responses. The researcher then further described the instructions given on the cover page of the test and turned the responsibility to supervise the survey to the class teacher. The students were given 40 minutes to answer and return the test to their teachers. However, only 165 students managed to complete the assessment.

TABLE 2. IL assessment domains and their respective items.

No	TRAILS-9 IL skills domains	Items
1	Develop, use, and revise search strategies (DU)	1–6
2	Develop topic (DT)	7–11
3	Identify potential sources (IP)	12–15
4	Evaluate sources and information (ES)	16–20
5	Recognize how to use information responsibly, ethically, and legally (RH)	21–25

TABLE 3. Person reliability.

	Raw score	Count	Measure	Model error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	11.3	25.0	-.29	.45	1.00	.0	1.00	.1
S.D.	3.7	.0	.75	.04	.12	.7	.20	.8
			SEPARATION 1.25			Person RELIABILITY .61		

Data analysis: Rasch model. A Rasch analysis for polytomous test items (test items or questions with three or more alternative responses, where only one is correctly scored or consistent with a targeted trait or other construct) was conducted using WINSTEPS® v. 3.68.2. In the context of the Item Response Theory (IRT), there are various models which characterize an item according to its “difficulty level” and “discrimination level.” Difficulty refers to the attribute of not being easily accomplished, solved, or comprehended. Meanwhile, item discrimination signifies the degree to which an item is differentiated among individuals with a higher or lower level of the trait or ability that is being measured (Cohen & Swerdlik, 2009).

First, it is important to ensure unidimensionality, that is, items in each IL domain assess a single construct. MNSQ (0.75–1.30) and Zstd (–2 to +2) fit statistics values determined that the response pattern for the items was unidimensional and the items were independent. Person and item reliability coefficients were calculated in order to assess classical test reliability (a priori alpha level was 0.05). An item-person map was then generated to explore the relationship between item difficulty and student ability.

The examination of person fit is essential in Rasch analysis because respondents with bizarre response patterns may seriously impact fit for the item level (Tennant & Conaghan, 2007). After the initial analysis, a total of 19 students (11.5%) were exempted from further analysis for being person misfit (careless, miscoding, or lucky-guessing).

Reliability. The Rasch model estimates a person’s reliability coefficient with the purpose of indicating the degree to which the test reliability orders test-takers based on their ability level. In the same manner, an item reliability index indicates the degree to which the items will remain in their difficulty order with repeated administrations (Salem, 2014). Both indices are placed on a 0 to 1 scale, with values closer to 1 sought because they are deemed desirable (Bond & Fox, 2007; Chan, Ismail, & Sumintono, 2014).

The summary statistics provided in Table 3 shows that the α value for person reliability is 0.61, which is moderately acceptable (Bond & Fox, 2007). However, the low person separation index (<2) indicates that the range of the students’ abilities is small, and it is very limited to consistently rank the ability of one student relative to another.

Summary statistics for the items are presented in Table 4. The value for item reliability is 0.94, which indicates that the items managed to create a well-defined variable and good reliability of item placement along the scale. This also implies that there are items that are difficult, moderate, and easy, which further suggest that the items are spread with a wide range of difficulty (Azrilah et al., 2013; Bond & Fox, 2007). A low item separation (< 2) and low item reliability (< 0.9) implies that the person sample is not large enough to confirm the item difficulty hierarchy (construct validity) of the instrument. The item separation index was 3.79, which indicates that item difficulty ranks will be consistent if the test is to be administered again.

The next step is to check for item reliability within each IL domain (construct) using the item separation index. Table 5 depicts that the items for the five IL domains possess item reliability ranging from 0.83 to 0.97, with ES having the lowest value. DU, DT, IP, and RH indices show that the items are high in reliability, with values close to 1.0 (Bond & Fox, 2007). This further indicates that the items are separated by a varying level of difficulty (Kamis et al., 2013). According to the result, it is revealed that the item separation index for each IL construct is between the values of 2.19 (moderate) to 5.15 (high), thus indicating a lack of item difficulty redundancy.

Person item map. The purpose of a measurement scale is that it is not too difficult or not difficult enough for the examinees. Rasch modeling is useful as it allows a linear representation of the items of the instrument to be placed on the same measurement scale as the person attributes. The person-item map, which is described as a graphical representation of person-abilities and item-difficulties, is drawn

TABLE 4. Item reliability.

	Raw score	Count	Measure	Model error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	65.8	146.0	.00	.19	1.00	-.1	1.00	-.1
S.D.	23.9	.0	.81	.01	.09	1.1	.13	1.1
			SEPARATION 3.79;			Item RELIABILITY .94		

TABLE 5. Item separation index.

No.	IL domains	Total items	Item reliability	
			Item	Separation*
1.	Develop, use and revise search strategies (DU)	6	0.91	3.22
2.	Develop topic (DT)	5	0.93	3.54
3.	Identify potential source (IP)	4	0.97	5.71
4.	Evaluate sources and information (ES)	5	0.83	2.20
5.	Recognize how to use information responsibly, ethically and legally (RH)	5	0.90	2.97

*Separation value more than 2 is good (Bond & Fox, 2007).

based on the equal measures (logits) of the raw item difficulties and raw person scores. The person's ability and item difficulty are shown in a straight line, which is also known as the variable map or Wright Map (Boone, Staver, & Yale, 2014). The items are placed based on the number of persons that manage to get a specific item correct. Persons are located based on the number of items that they are able to answer correctly (Wright & Stone, 1999, p. 151).

The person-item map is divided into four quadrants. Person estimates are distributed on the left side, while item estimates are on the right side based on their ability and difficulty estimates (Tee et al., 2013). The students in the upper left are said to be "better" or "smarter" than the items on the lower right, which implies that the easier items are not difficult enough to challenge the highly proficient students. On the other hand, the items on the upper right outsmart students on the lower left, which further suggests that the difficult items are beyond their ability level.

Figure 1 illustrates the person-ability, item-difficulty map in a consistent spread. The items and persons are spread evenly along the standardized scale and are clustered opposite each other. The spread of item measure ranges from a maximum logit value of 1.50 and minimum logit value of -1.33 (left side of the figure). The most difficult item (IP13: *characteristics of a primary information source*) remains at the top right, while the easiest item (DU4: *use of truncation in Google search*) is placed in the lower right of the scale. Highly capable students managed to answer the difficult items, while students with lesser ability failed to provide the right answer. If a student is plotted at the same level as an item, it means that the student has a 50% chance of responding to that item correctly.

There are no significant gaps in terms of difficulty between items. Item IP13 clearly falls above the ability of almost all students (difficult item), which may be potentially eliminated from the assessment because it does not contribute to the precision of measurement. Nevertheless, all other items fall within the ability of the students. There are no serious gaps across the items or student ability, which further indicates that the test is able to distinguish the range of student abilities fairly effectively and it is well targeted for this group. The assessment also included both difficult and easy items, as evident in the spread of items along the standardized linear scale. Although the clumping of five items (located in a row at about -0.3 logit) imply that the items

are psychometrically redundant relative to their difficulty, this is acceptable because they belong to a different IL skills domain. The general pattern shows that more students are plotted below the item mean at 0.00 logit. The abilities of the students in developing topic (DT) and develop, use, and revise search strategies (DU) are better than their ability to identify potential sources (IP) and evaluate sources and information (ES). RH (ethical, responsible, and legal use of information) is also found to have a lower difficulty level based on student ability. The map is evenly represented considering that every item in each category moves from an easier to a more complex ability.

Discussion

The present study provides an introduction to the use of Rasch modeling in validating the information literacy assessment instrument. The advantage of using Rasch modeling over the classic test theory (CTT) refers to its ability to provide an item function-based analysis that allows a closer look at how each item contributes to the overall assessment. The application of the Rasch analysis revealed several important facts. First, the analysis showed that the test was unidimensional, which indicates that the items in each domain only measured their respective skill. Second, item difficulty was not associated with item domain, whereby the items in each domain of IL ranged from less difficult to very difficult. However, items related to identifying potential sources (IP) and evaluating sources and information (ES) were found to be more difficult compared to developing topic (DT) and develop, use, and revise search strategies (DU) as well as ethical, responsible, and legal use of information (RH). Third, the test demonstrated that students in general failed to attain a high level of competence. Moreover, the test managed to easily differentiate between students of high ability and low ability. Hence, it could be used to rank one student as having more ability than the other because none of the students had the ability above the items—which implies the high odds of a student to respond to any item correctly.

Nevertheless, no evident problem was found with the test items, except for item IP13, with only 16% of students who managed to answer this item correctly. The purpose of this item is to ask the students to identify the main characteristics of a primary source of information. It is surprising that very

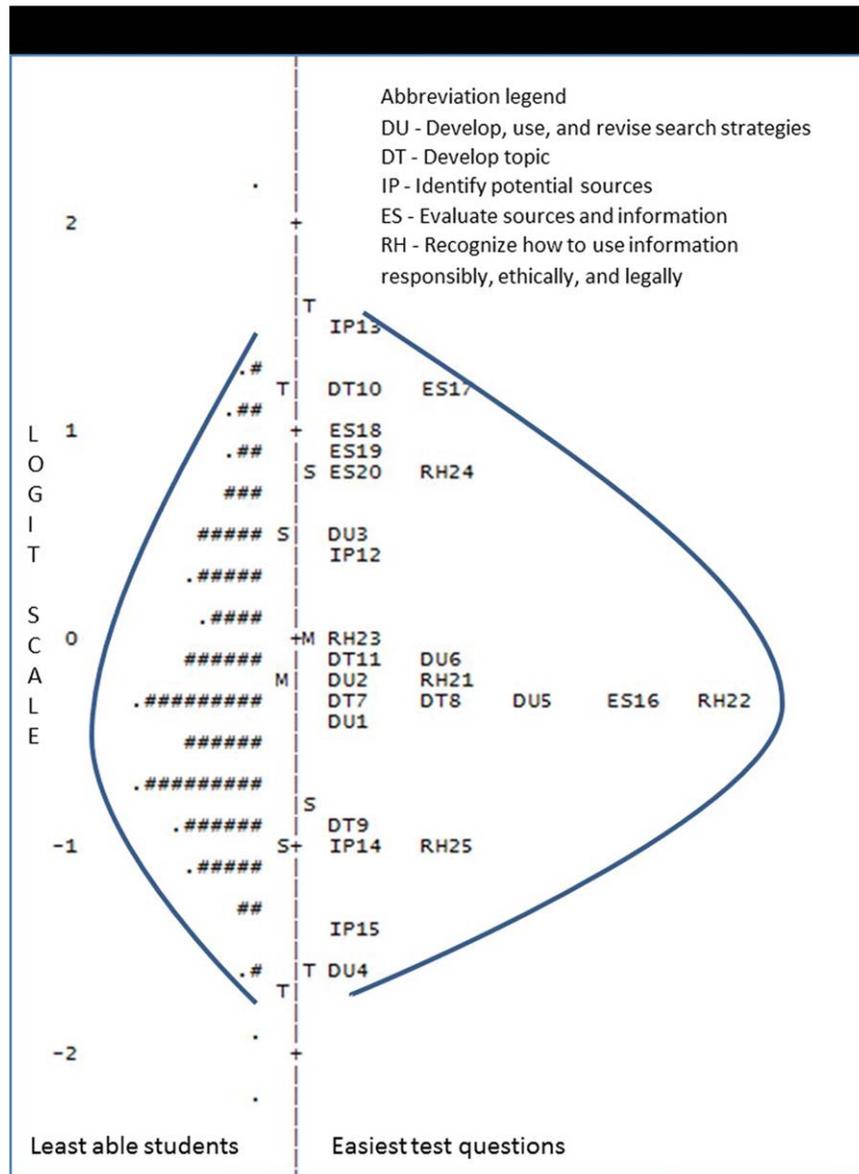


FIG. 1. Person-item map. [Color figure can be viewed at wileyonlinelibrary.com]

few students could answer it correctly during the actual assessment considering that there were no problems with this question during the pilot test. A possible explanation would be that the students were confused with the wording of the choices given. It should be taken into consideration that complex wording may be regarded as a test of reading comprehension instead of an assessment of whether the student knows the subject matter. Hence, the item is suggested to be changed in order to improve the assessment. The results of the map also indicate that there is a possibility that additional items will be included in the tool, but with different levels of difficulty. The clustering of several items along logit 0.0 to -0.4 and a gap between logit -0.4 to -1.0 can be further improved by reassessing the item difficulty level.

In the category of Developing Topic [DT], items DT7, DT8, and DT11 were developed based on almost equal

difficulty level, while item DT10 is of a much higher difficulty level considering that only 22% managed to answer it correctly. Both DT7 and DT8 required the students to identify whether the topic is “broad” or “narrow” and is of the same difficulty level. It is suggested that one of the items be dropped or replaced with another item for testing DU skills. The explanation for the difficulty level of item DT10 refers to the negativity of the question. In relation to this, students were asked to select which of the multiple-choice answers does not represent a general to specific topic sequence. Hence, the item should be reworded positively to identify the sequence ranging from a general topic to a specific topic.

Items measuring DU (develop, use, and revise search strategies) comprised a wider range of item difficulty distribution, but with no overlap of item difficulty. Most of the

students did not report any difficulty with this domain. Specifically, item DU4 is labeled as the least difficult, while DU1, DU2, DU5, and DU6 with average difficulty. In relation to this, only DU3 was considered a difficult item because it requires the students to select the correct research steps (information-seeking process) in planning a menu (information need). However, only 34% of the students answered this item correctly. Nevertheless, it is important to note that this item was a direct translation from TRAILS-9 without any modification. The easiest item was DU4, which provides the truncated search term (*child\$or child**) which requires the student to identify words retrieved from a Google search. Surprisingly, a total of 77% students managed to answer this correctly despite the skepticism shown by the teachers regarding the use of truncation. As a result, this item is suggested to be retained. DU1, DU2, DU5, and DU6 were found to sit closely on the map but with no overlapping. Given that the separation index is 3.22, this is acceptable.

Based on the findings, the domain that may need reassessing is the ES. On the map, ES18, ES19, and ES20 respectively representing recognizing, authority, and biasness and credibility are closely located, thus indicating an almost similar level of difficulty. Only a small number of students were able to answer these items correctly. Hence, it can be concluded that too many difficult items may not be suitable for the students' ability. ES17 (currency of a website) is a highly difficult item, with only 22% of the students answering it correctly. Meanwhile, ES17 and ES18 (25% answered correctly) are negatively worded items. This makes it difficult for students to eliminate correct answers, especially when the distractors were too close to the correct answer.

Test items relating to RH (ethical, responsible, and legal use of information) were found to be fairly distributed in terms of difficulty related to the students' ability. The spread along the line is acceptable at an item separation index of 2.27. However, only item RH24 was proven to be more difficult, as it only managed to obtain 28% correct answers. In this case, the students did not have the ability to identify the correct form of paraphrasing. Nevertheless, most students could identify a book's title when an image of the book's title page is provided to prepare a bibliography (RH25). Overall, all items of RH are suggested to be retained.

Additionally, the overall performance level of the students was average, which is not surprising because these students have no formal IL education. They are only exposed to several IL related skills as the teachers prepare them for their project work and reports. In general, it was observed that even the lower ability students were found to be somewhat competent to use the information responsibly and ethically. On top of that, they also possess the ability to develop a topic as well as to devise search strategies. However, most students have limited ability to identify and evaluate information sources.

More important, these findings provide implications for test developers. The difficulty of an item can vary depending

on the associated IL domain. Moreover, it cannot be assumed that developing a topic is an easier IL skill compared to identifying potential sources or devising a search strategy. In other words, each skill set requires its own level of competence. The second important fact is that this test cannot be used as a mastery level test, with the assumption that the students have the ability to correctly answer most items. In this case, the TRAILS-9 test demonstrated that the items are above the competency level of the average student's ability. Hence, this test can be used by teachers to identify difficult skills as well as work on increasing students' competency in the related skill.

The assumption made in this study was that the respondents provided sincere responses to the test items. There were no formal IL skills sessions given to the students; therefore, they had to rely on their own competencies to answer the test items as best they could. This study was comprised of a single test that was administered on a single group of students at one school; hence, it consequently limits the generalizability of the findings. However, the methods used in developing the adapted and translated assessment tool as well as the validation of the students' performance through Rasch modeling can be applied to other tests and cross-cultural studies on IL.

The standard error of calibration for each individual item as well as the standard error of measurement for each person's ability was obtained using Rasch calibration. With traditional methods, a standard error of measurement is provided only for measures at the group mean of person ability (Wright & Stone, 1999, p. 155).

Conclusion

Information literacy (IL) is a soft-discipline that draws upon theory and research approaches obtained from other disciplines, particularly educational theories and research approaches (Johnston & Webber, 2006). There is a need to develop the research base and nurture knowledgeable LIS researchers if LIS is to be established as a discipline. The growth of IL research from educational base to workplace IL (Jinadu & Kiran, 2014, 2016) is evident, but there has not been a marked development in the IL assessment methodology. Hernon and Schwartz (2016) in their editorial piece on "making connections" imply the need to draw linkage from a particular study to something of broader relevance, including the use of a methodology that is not common in LIS research. Hence, this study attempted to establish a guideline for a systematic adaptation, translation, and validation of an information literacy assessment scale in a possible cross-cultural context. There have been numerous isolated efforts to develop scales or tests to assess information literacy but the duplication of efforts is a waste of resources. The reuse of preexisting test instruments for the purpose of adaptation and translation strategies not only retains the original conceptualization of the test items, but also ensures the accuracy even though it is performed under different cultural contexts.

The present study has demonstrated that the adapted and translated TRAILS can be used as a valid and reliable instrument for secondary school students' information literacy assessment. The adaptation of a preexisting test instrument involves:

1. Identification of test content: decide which performance standard is the most suitable for the educational context.
2. Item translation: involves certified translators in both the original language and target language as well as subject experts such as teachers and librarians. The translators focus on colloquial phrases, slang, and jargon, while the teachers and librarians work together to ensure the items retain their conceptual and content accuracy as well as being comprehensible to the students. Meticulous scrutiny of each item is necessary to ensure that the ability of the students to answer correctly is definitely based on their competence instead of a poor design of the question. The reuse of assessment scales with documented adaptations is expected to contribute to the universal IL assessment to some degree.
3. Test reliability: construct validity and scoring interpretations when using an IL assessment test. The validation of the adapted test instrument using Rasch modeling serves as a tool to assist educators and students to accurately assess performance. Student abilities are matched with the difficulty levels of the test items. This provides further information to allow improvements to be made to the assessment instruments, including the areas of improvement based on students' competence in the skill.

The current research has contributed to the field of measurement. First, researchers interested in adapting a preexisting scale in a cross-cultural context will now have clear guidelines on the translation process to ensure the validity of the test items within a new context. Second, each test item can be mapped onto students' ability using the Rasch analysis method, thus ensuring the reliability of the test outcomes. Items can be modified or adjusted according to their difficulty level, while the IL proficiency of the students can be assessed based on item difficulty level instead of only referring to student ability level. Pinto (2015) warns that excessive clusters and overlapping among fields in IL assessment research show that the IL subdomains are "still being developed," which further indicates the need for a more consolidated field. In this study, the Rasch analysis was conducted on a small sample for a preliminary assessment of items' psychometric properties. Hambleton (2005) believed that adapting an instrument to compare data from different samples and from different backgrounds enables a greater ability to generalize and investigate differences within an increasingly diverse community or population. It is hoped that cross-cultural research can help IL educators understand the complexity of IL assessment based on the same theoretical and conceptual foundations. Overall, this understanding will help build IL as a discipline with a strong methodological foundation.

Acknowledgment

The present study was financially supported by the Ministry of Education Malaysia.

References

- Al-Aufi, A., & Al-Azri, H. (2013). Information literacy in Oman's higher education: A descriptive-inferential approach. *Journal of Librarianship and Information Science*, 45(4), 335–346.
- American Association of School Librarians and Association for Educational Communications and Technology. (1998). *Information Literacy Standards for Student Learning*. Chicago: American Library Association. Retrieved from http://umanitoba.ca/libraries/units/education/media/InformationLiteracyStandards_final.pdf
- Association of College and Research Libraries (ACRL). (2000). *Information literacy competency standards for higher education*. Chicago: American Library Association. Retrieved from <http://www.ala.org/acrl/standards/informationliteracycompetency>
- Association of College and Research Libraries (ACRL). (2015). *Document ACRL MW15 Doc 4.0*. Retrieved from <http://acrl.ala.org/ilstandards/wpcontent/uploads/2015/01/Framework-MW15-Board-Docs.pdf>
- Azrilah, A. A., Mohd Saidudin, M., Faridah, M.I., & Mohd Zaidi, O. (2013). Insights into engineering education learning outcome's assessment with Rasch model. *Research Journal of Applied Sciences, Engineering and Technology*, 6(19), 3520–3526. Retrieved from <http://maxwellsci.com/print/rjaset/v6-3520-3526.pdf>
- Baji, F., & Bigdeli, Z. 2016. Adapting and normalizing the 6th grade version of the Tool for Real-Time Assessment of Information Literacy Skills (TRAILS) among the Iranian 6th grade students. *Pakistan Journal of Information Management & Libraries*, 17, 137–145.
- Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). *Guidelines for the process of cross-cultural adaptation of self-report measures*. *Spine*, 25(24), 3186–3191.
- Boh Podgornik, B., Dolničar, D., Šorgo, A., & Bartol, T. (2015). Development, testing, and validation of an information literacy test (ILT) for higher education. *Journal of the Association for Information Science and Technology* (in press). doi: 10.1002/asi.23586.
- Bond, T.G., & Fox, C.M. (2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). New York: Routledge.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht, Netherlands: Springer.
- Cameron, L., Wise, S. L., & Lottridge, S. M. (2007). The development and validation of the information literacy test. *College & Research Libraries*, 68(3), 229–237.
- Chan, S.W., Ismail, Z., & Sumintono, B. (2014). A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. *Procedia Social and Behavioral Sciences*, 129, 133–139.
- Chang, Y., et al. (2012). Assessing students' information literacy skills in two secondary schools in Singapore. *Journal of Information Literacy*, 6(2), 19–34.
- Chu, S.K.W. (2012). Assessing information literacy: A case study of primary 5 students in Hong Kong. *School Library Research*, 15, 1–24. Retrieved from <http://web.hku.hk/~samchu/docs/Chu-2012-Assessing-information-literacy-A-case-study-of-primary-5-students-in-Hong-Kong.pdf>
- Cohen, R.J., & Swerdlik, M.E. (2009). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (7th ed.). New York: McGraw Hill.
- Diller, K.R., & Phelps, S.F. (2008). Learning outcomes, portfolios, and rubrics, oh my! Authentic assessment of an information literacy program. *Portal: Libraries and the Academy*, 8(1), 75–89.
- Dorner, D.G., & Gorman, G.E. (2006). Information literacy education in Asian developing countries: cultural factors affecting curriculum development and programme delivery. *IFLA Journal*, 32(4), 281–293.
- Foo, S., et al. (2013). Information literacy skills of humanities, arts, and social science tertiary students in Singapore. *Reference & User Services Quarterly*, 53(1), 40–50.

- Foo, S., et al. (2014). Information literacy skills of secondary school students in Singapore. *Aslib Journal of Information Management*, 66(1), 54–76.
- Gall, M.D., Gall, J.P., & Borg, W.R. (2003). *Educational Research: An Introduction* (7th ed.). Boston: A & B Publications.
- Gola, C.H., Ke, I., Creelman, K.M., & Vaillancourt, S.P. (2014). Developing an information literacy assessment rubrics: A case study of collaboration, process and outcomes. *Communications in Information Literacy*, 8(1), 131–144. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.682.4549&rep=rep1&type=pdf>
- Gross, M., & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library & Information Science Research*, 29(3), 332–353.
- Gross, M., & Latham, D. (2012). What's skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology*, 63(3), 574–583.
- Gross, M., & Latham, D. (2013). Addressing below proficient information literacy skills: Evaluating the efficacy of an evidence-based educational intervention. *Library & Information Science Research*, 35(3), 181–190.
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46(12), 141.
- Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R.K. Hambleton, P.F. Marendia, & C.D. Spielberg (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum.
- Heron, P., & Schwartz, C. (2016). Making connections. *Library & Information Science Research*, 38(1).
- Jinadu, I., & Kiran, K. (2014). Information literacy at the workplace: A suggested model for a developing country. *International Journal of Libraries and Information Services*, 64(1), 61–74.
- Jinadu, I., & Kiran, K. (2016). Practices leading to information literacy development among Nigerian undergraduates. *Malaysian Journal of Library & Information Science*, 21(1), 109–121.
- Johnston, B., & Webber, S. (2006). As we may think: Information literacy as a discipline for the information age. *Research Strategies*, 20, 108–121.
- Kamis, A., Bakar, A.R., Hamzah, R., Asimiran, S.A., & Halim, N.A. (2013). Competency assessment of clothing fashion design: Rasch measurement model for construct validity and reliability. *Journal of Technical Education and Training (JTET)*, 5(2), 1–13.
- Korb, K.A. (2012). Conducting educational research: Steps in conducting a research study. Retrieved from <http://korbedpsych.com/R09aAdopt.html>
- Korobili, S., Malliari, A., & Christodoulou, G. N. (2009). Assessing information literacy skills in the Technological Education Institute of Thessaloniki, Greece. *Reference Services Review*, 37(3), 340–354.
- Kovalik, C.L., Yutzey, S. D., & Piazza, L.M. (2012). Assessing change in high school student information literacy using the tool for real-time assessment of information literacy skills. *Contemporary Issues in Education Research*, 5(3), 153–166. Retrieved from <http://www.cluteonline.com/journals/index.php/CIER/article/view/7092>
- Kurbanoglu, S.S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation*, 62(6), 730–743.
- Latham, D., & Gross, M. (2008). Broken links: Undergraduates look back on their experiences with information literacy in K-12 education. *School Library Media Research*, 11.
- Majid, S., Chang, Y.K., & Foo, S. (2016). Auditing information literacy skills of secondary school students in Singapore. *Journal of Information Literacy*, 10(1), 44–66.
- Miller, C. (2016). TRAILS: Tool for Real-time Assessment of Information Literacy Skills. *The Charleston Advisor*, 17(3), 43–48.
- Ngo, H., Walton, G., & Pickard, A. (2016). An assessment of the information literacy capability of Vietnamese upper secondary students. In *Proceedings of the Western Balkan Information Literacy Conference*. Pp. 65–70. Retrieved from https://blog.stcloudstate.edu/ims/files/2016/06/Proceedings_WBILC2016-25s3dpu.pdf#page=66
- Oakleaf, M. (2008). Dangers and opportunities: a conceptual map of information literacy assessment approaches. *Portal: Libraries and the Academy*, 8(3), 233–253.
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science & Technology*, 60(5), 969–983.
- Pinto, M. (2015). Viewing and exploring the subject area of information literacy assessment in higher education (2000–2011). *Scientometrics*, 102(1), 227–245.
- Rattanaowongsa, R., & Koraneekij, P. (2015). A development of blended information literacy learning web for Thai high school students. *Procedia Social and Behavioral Sciences*, 174, 2693–2699
- SAILS. (2000). *Project SAILS (Standardized Assessment of Information Literacy Skills)*. Kent, OH: Kent State University.
- Salem, J. (2014). *The Development and Validation of All Four TRAILS (Tool for Real-Time Assessment of Information Literacy Skills) Tests for K-12 Students*. (Ph.D. Dissertation). Kent, OH: Kent State University.
- Schloman, B.F., & Gedeon, J.A. (2007). Creating TRAILS: Tool for Real-Time Assessment of Information Literacy Skills. *Knowledge Quest*, 35(5), 44–47.
- SCONUL. (2012). *The SCONUL Seven Pillars of Information Literacy*. Retrieved from <http://www.informationliteracy.org.uk/2012/01/sconul-seven-pillars-of-information-literacy-new-lens-now-available/>
- Smith, J. K., Given, L. M., Julien, H., Ouellette, D., & DeLong, K. (2013). Information literacy proficiency: Assessing the gap in high school students' readiness for undergraduate academic work. *Library & Information Science Research*, 35(2), 88–96.
- Sousa, V.D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274.
- Tee, O.P., Law, N., Soojin, K., Kim, S., & Ki, T.S. (2013). Psychometric assessment of ICILS test items on Hong Kong and Korean students — A Rasch analysis. In *The 5th IEA International Research Conference*. Singapore. Retrieved from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC-2013_Oon_et_al.pdf
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358–1362.
- TRAILS. (2004). *Tool for Real-time Assessment of Information Literacy Skills*. Kent State University Libraries. Retrieved from <http://www.trails-9.org>
- van Helvoort, A.A.J. (2013). How faculty in the Hague University of Applied Sciences uses the scoring rubric for information literacy. In *Worldwide Commonalities and Challenges in Information Literacy Research and Practice* (pp. 436–442). Berlin: Springer International Publishing.
- Walsh, A. (2009). Information literacy assessment: Where do we start? *Journal of Librarianship and Information Science*, 41(1), 19–28.
- Whitlock, B., & Ebrahimi, N. (2016). Beyond the library: Using multiple, mixed measures simultaneously in a college-wide assessment of Information Literacy. *College & Research Libraries*, 77(2), 236–262.
- Willer, D.B. (2014). *Targeting Success. An Evaluation of Information Literacy Standards: A Mixed Method Approach Utilizing the Judgments of National Board Certified Teachers*. PhD Dissertation. Seattle, WA: University of Washington.
- Wright, B.D., & Stone, M.H. (1999). *Measurement Essentials* (2nd ed.). Wilmington, DE: Wide Range.
- Yu, H., Abrizah, A., & Sani, M. K. J. A. (2016). Information literacy through resource-based learning: Malaysian teachers' conception and instructional practices. *Malaysian Journal of Library & Information Science*, 21(1), 53–67.