ASSESSING THE VALIDITY AND RELIABILITY OF THE INVENTORY OF SCIENCE TEACHER READINESS IN IMPLEMENTING CLASSROOM-BASED ASSESSMENT (ISTRI-CBA)

***Mat Rasid Ishak[1]**
Hidayah Mohd Fadzil[1]
Harris Shah Abd Hamid[2]
[1] Department of Mathematics and Science Education,
Faculty of Education, Universiti Malaya
[2] Faculty of Management, Education and Humanities, University College MAIWP International, Kuala Lumpur, Malaysia
**matrasid.ishak@gmail.com*

ABSTRACT

Teacher readiness is the willingness of a teacher to implement a planned program successfully. Readiness is crucial in ensuring a program can be implemented at the individual or organizational level. The readiness of science teachers to implement Classroom-Based Assessment (CBA) determines the direction and success of primary school assessments in Malaysia. This pilot study aims to create and evaluate the Inventory of Science Teachers' Readiness in Implementing Classroom-based Assessment (ISTRI-CBA). There have been many studies on CBA. However, there are limitations in measuring the readiness of science teachers to implement CBA. The evaluation will be ineffective without a suitable tool to measure teacher readiness in CBA. Therefore, to determine the level of readiness of science teachers, the researcher needs an instrument that is valid, reliable, and appropriate to the context of the Ministry of Education, culture, and the current situation in Malaysia. It is necessary to have a conceptually and psychometrically substantial inventory. This challenge was addressed by developing the ISTRI-CBA, comprising five measurement dimensions: knowledge about CBA, skills of CBA, resource support, attitudes, and professional values. The development of ISTRI-CBA requires validity and reliability for its use. This study aims to (1) test the quality characteristics of the ISTRI-CBA items and (2) determine the reliability of the ISTRI-CBA obtained using the Rasch Model measurement. The research involves the administration of a survey questionnaire to 44 science teachers in a district in Malaysia. The findings show that the ISTRI-CBA, comprising 157 items, has good psychometric characteristics and meets the measurement criteria of the Rasch measurement model. The item quality and reliability also prove the suitability of the whole dimension. The study's findings provide evidence of the empirical implications that consistently support ISTRI-CBA as a valid and reliable instrument to measure teachers' readiness to implement CBA in science subjects. The findings also reveal that the ISTRI-CBA can measure science teachers' readiness to implement CBA.

Keywords: *Classroom-Based Assessment, Science Teacher Readiness, Rasch Measurement Model, Validity*

INTRODUCTION

Classroom-based Assessment (CBA) in science subjects in Malaysia has been introduced since 2019 (Curriculum Development Department, 2019). Teachers' readiness to implement CBA is the key to successful implementation for years to come. Research on readiness began in the 2000s, and scientists began to study readiness in individuals through experience, environment, and affective (Madsen et al., 2006). These individual readiness factors contribute to the ability to make the right decisions in implementing policy changes to increase the effectiveness of their implementation. Readiness studies have focused on the factors of individual readiness to encourage change as the first step in the successful implementation of a program (Khalil Fakhri Abdul & Mohd Isha Awang, 2016). Recent research links individual or teacher readiness with practical findings such as readiness for assessment, learning, and teaching skills (Arar & Abramovitz, 2017); Bouillet et al., 2014); Kakia et al., 2016; Moosa et al.,2020a); achievement performance (Lynch et al.,2019), work performance, training and consideration of professionalism (Avidov-Ungar & Arviv-Elyashiv, 2020) have encouraged the development of teacher competence and the feasibility of change in various fields, especially assessment classroom.

Although the effect of individual readiness is gaining attention, the impact of interventions related to teacher readiness to implement CBA in science is challenging to measure, mainly due to the lack of a universally accepted definition for measuring teacher readiness to implement CBA in science (Christensen & Knezek, 2017). There are several instruments for measurement in assessment; among those instruments are teacher readiness in assessment by Rorlinda Yusof & Faridah Anum Abdul Wahid (2014); Khalil Fakhri Abdul & Awang Mohd Isha (2016); Arsaythamby Veloo et al. (2015); Abdul Said Ambotang & Shanti Gobalakrishnan (2017), an instrument to measure the level of pre-service teacher assessment readiness and knowledge (Zahari Suppian, 2014), an assessment Literacy Model instrument (Rohaya Talib & Mohd Najib Abd Ghafar, 2008) and an Inventory Teacher Assessment Practice (Suah et al., 2010).

Indeed, no study measures the readiness of teachers to implement CBA in science subjects. Research conducted to develop an instrument that can accurately measure according to the operational definition in the proper context is necessary to obtain validity and reliability in data analysis. Although many validation studies have been conducted on the instrument, the generalizability of any instrument from other non-Malaysian countries is still unclear according to the appropriate content and context. However, the study of findings conducted by Bruce (2018b) has emphasized the consequences of not having valid and reliable instruments for measuring CBA in Malaysia, causing the instruments used to measure the level of preparedness of science teachers to be inaccurate and not generalized.

Therefore, the main objective of this research is to extend the literature on new instruments for measuring teachers' readiness to implement CBA in science subjects by evaluating psychometric properties using the Rasch Model Measurement based on a sufficient sample of 44 science teachers (John, 1994). This finding can be used for further research in Malaysia and in Asia in general to measure the readiness of teachers to implement CBA in science subjects.

STUDY OBJECTIVES

This study aims to develop an inventory to measure the readiness of science teachers to carry out classroom assessments in public primary schools. Two specific objectives of the study are determined to achieve the desired goal:

1. To test the quality characteristics of ISTRI-CBA items using the Rasch Model.
2. To determine the reliability of ISTRI-CBA using the Rasch Model.

LITERATURE REVIEW

Defining Teacher Readiness

Readiness refers to the extent to which a teacher or individual is ready to understand and deepen a given program (Kang et al., 2019). In the context of this study, readiness is defined as the science teacher's ability to prepare himself to implement CBA from the aspects of learning, environment, and affective aspects of the implemented program. The readiness of science teachers in this study was measured using a construct translated from a survey of previous studies related to the readiness of teachers to carry out assessments. The science teacher readiness construct used in ISTRI-CBA consists of the construct of CBA knowledge, CBA implementation skills, resource support provided in implementing CBA, science teacher attitudes, and professionalism values.

Classroom-based Assessment Programmes

Classroom-based Assessment (CBA) is an assessment carried out to obtain information about student development, student abilities, and current student mastery. In its book, the Curriculum Development Division (2019) translates CBA through two types of assessment, namely formative and summative. Both of these assessments are done continuously in the classroom to ensure student development. Teachers have a crucial role in the implementation of CBA through the setting of learning objectives based on content standards and assessed learning standards. Teachers need to plan, build instruments, implement, record, analyze, report, and follow up on student assessments. All these constituents need to be implemented by teachers in order to collect and analyze information continuously so that teachers can make judgments to improve the assessment carried out in the classroom (Ministry of Education Malaysia, 2017a). In this study, the focus is on measuring the willingness of science teachers to implement CBA through an instrument developed by the researcher.

Classroom assessment in science subjects is a continuous assessment in T&L to obtain information related to student progress, abilities, and achievements (Mohd Haidzir Yusof@Jusoh& Norasmah Othman, 2019). CBA in science subjects involves T&L, which includes the assessment of knowledge, skills, and values. CBA assessment is conducted with various activities such as simple projects, quizzes, games, role-playing, and storytelling that replace the mid- and end-of-year centralized exams. The student's work done in the form of a project will be evaluated as exam assessment material, which will be translated into the student's Proficiency Level (PL). The student's PL in knowledge, scientific skills, attitudes, and moral values as a whole is done by using the value of the teacher's professional judgment (Curriculum Development Department, 2019). The overall PL definition is divided into PL 1 to PL 6, as shown in Table 1 below.

Table 1. KSSR Elementary School Science General Proficiency Level Statement

Proficiency Level	Interpretation
1	Students are **knowledgeable** on the basics related to science knowledge and skills as well as show interest in teaching and learning of science or can implement the basic skills of science concepts or respond well in the basics related to science knowledge for a topic taught in class.
2	Students can respond and **understand** science knowledge and skills as well as show an understanding of the science taught by being able to explain a topic orally while showing a deep curiosity.
3	Students can **apply** science knowledge and skills learned in class to perform simple tasks honestly and are able to record science data accurately in tasks or situation in the classroom.
4	Students can **analyze** science knowledge and skills learned in teaching and learning in the context of systematic problem solving and bravery in trying new things.
5	Students can **evaluate** science knowledge and skills to implement science process skills on assignments for problem solving situations and make decisions according to procedures in a systematic, analytical, collaborative, diligent and persevering manner in the classroom.
6	Students are able to **create** in relation to science knowledge and skills in the context of problem solving and make decisions in a variety of situations creatively and innovatively in order to generate new ideas. Students also have exemplary qualities, responsible for themselves, friends, the environment and highly prudent.

Khalil Fakhri Abdul and Awang Mohd Isha (2016) stated that the importance of science teachers to appreciate and understand related knowledge and skills emphasized in the implementation of assessment is a factor in the effectiveness of the implementation of the CBA program. The effectiveness of assessment is also in line with the training obtained by science teachers in CBA and is further linked to student success (Koloi-Keaikitse, 2016). In this regard, teachers' readiness to understand and appreciate how to implement CBA in science subjects is essential because it can ensure that the assessment can produce students as desired in the National Education Philosophy.

Theoretical Framework

This study adopted the Theory of Individual Readiness (Madsen et al., 2006). This theory has been combined with five Assessment Models to determine the construct of science teacher readiness in CBA, namely Teacher Concept of Assessment (Brown, 2004), Educational Assessment Knowledge and Skills for Teachers (Brookhart, 2011), Development and Validation of the Attitude Toward Educational Measurement Inventory (Bryant & Barnes, 1997), Assessment Literacy Model (Rohaya Talib & Mohd Najib Abd Ghafar, 2008) and Model Assessment Teacher Practice Inventory (Suah et al., 2010). The result of this adaptation of the construct for the readiness of science teachers in assessment is depicted in Figure 1.

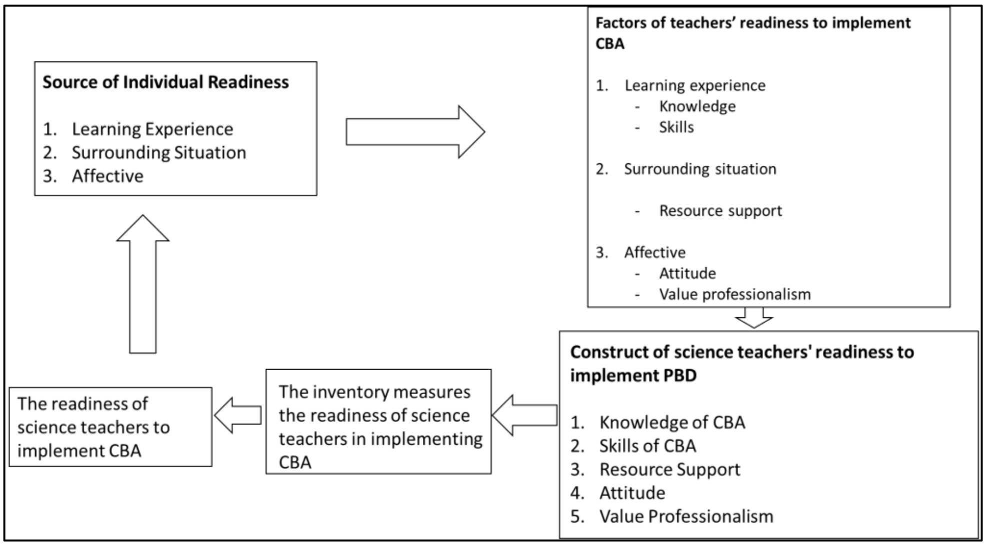

Figure 1. Theoretical Framework of the Construct of Science Teacher Readiness

Individual Readiness Theory by Madsen et al. (2006) which was adopted, has elements of individual readiness for a teacher: learning experience, environmental, and affective factors. Factors of learning experience, environmental, and affective conditions, when referred to the analysis results based on five assessment models, found that the construct of teacher readiness in assessment for learning experience consists of teacher knowledge and skills in CBA. The environmental factor consists of resource support systems in infrastructure facilities and structures, while the affective factor consists of attitudes and values of professionalism.

Conceptual Framework

The conceptual framework of the study in Figure 2 is adapted from the Four Building Blocks Theory Instrument Construction Model (Four Building Blocks Theory) by Wilson (2005), Individual Readiness Theory by Madsen et al. (2006), a dimensional analysis of the concept of science teachers' readiness in assessment as well as an analysis of five assessment models. The first block in Figure 2 shows the mapping of the science teacher readiness construct, translated into sub-constructs that comprise the Science Teacher Readiness Inventory for Classroom Assessment (ISTRI-CBA) items.

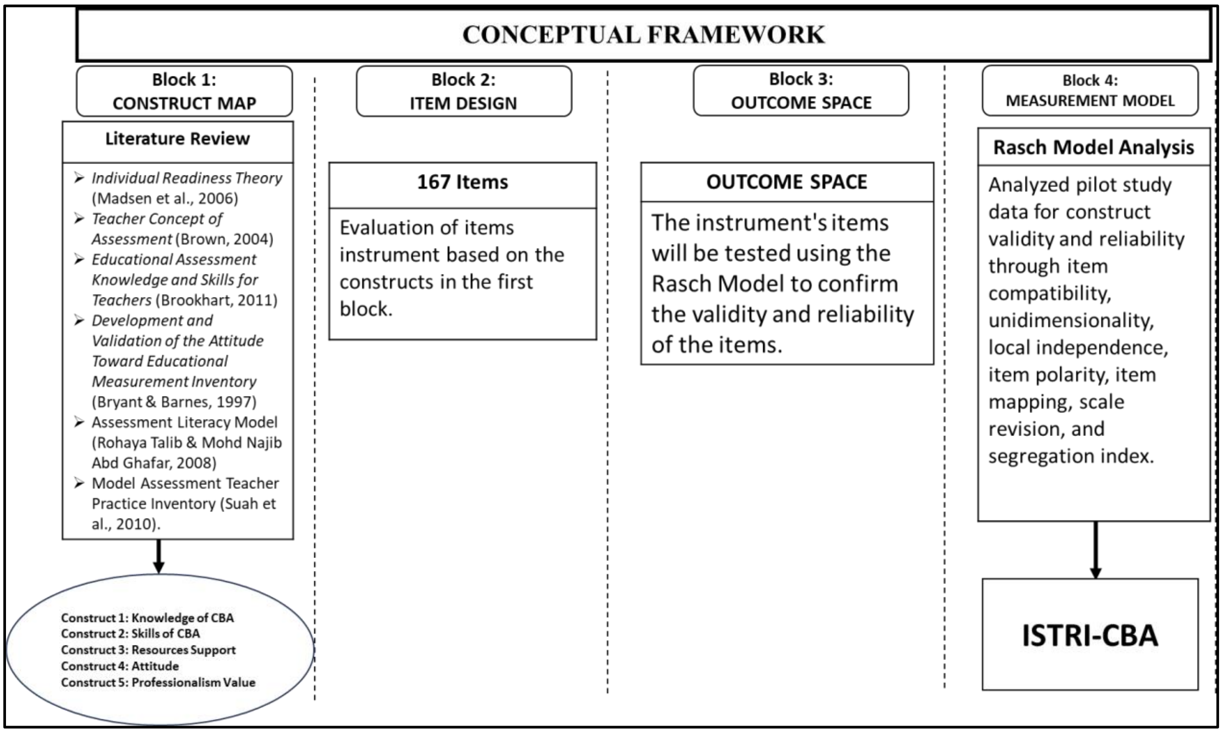

Figure 2. Conceptual Framework of ISTRI-CBA

The first block in Figure 2 above shows five dependent variables known as the construct of science teachers' willingness to implement CBA, which consists of CBA knowledge, CBA skills, resource support, attitudes, and professionalism values. These five variables are placed in an ellipse, which illustrates that each variable will have an independent variable represented by a sub-construct that can be measured through items developed by the researcher. The second block in the conceptual framework of this study involves the selection of ISTRI-CBA items with content validity through expert panel review findings. A total of 167 best items with confirmed contents were accepted and used for this pilot study. Next, in the third block, which is the results room, the instrument's items were tested using the Rasch Model to confirm the validity and reliability of the items. In the last phase, which is the fourth block, testing, and analysis were done to determine the appropriate psychometric characteristics in the Rasch analysis, which aims to determine the selection of the final ISTRI-CBA items that are developed to measure the construct of teacher readiness in CBA (Rasch, 1993). The validity and reliability of the items obtained in this pilot study ensure that all the ISTRI-CBA items have been prepared for the following study, which is a validation study. Therefore, the psychometric characteristics of the items studied in ISTRI-CBA were tested using the Rasch Model to obtain the validity and reliability of the inventory with an analysis of item compatibility and unidimensionality first to ensure that the items meet the Rasch measurement model. Then, the test continued with the analysis of local independence, item polarity, individual item mapping, scale review, and segregation index (Rasch, 1993).

METHODOLOGY

Participants

A total of 44 teachers who teach science subjects in schools around the Hulu Langat District participated in this study. The sample consisted of 15 national schools, a Chinese National School, and a Tamil National School covering urban and rural areas. The study respondents consisted of science teachers with teaching experience from 3 months to 35 years of teaching who were classified into new teachers (3 months to 10 years), intermediate teachers (11 to 20 years) and experienced teachers (21 to 35 years of teaching).

Measure
The study used the ISTRI-CBA Instrument, which consists of a four-point Likert-rating scale (see Supplementary Table in the Appendix). A four-point Likert-rating scale ranges from 1 (Strongly Disagree) to 4 (Strongly Agree).

Procedure
The research was approved by the Malaysian Ministry of Education (Reference No.: KPM 600-3/2/3-eras (13026) before data was collected between July and August 2022. ISTRI-CBA is one of the steps in a new study to identify teachers' readiness to implement CBA in science subjects for primary schools. A total of 44 out of 45 science teachers for this study have provided feedback in this study, and the response rate is 97.78%.

Data Analysis
Rasch analysis was performed on pilot study data using Winsteps (Version 5.1.0; Linacre, 2019b). The Rasch Rating Scale Model (Andrich, 1978) was used in this study, given its suitability for measurement with items that share the same rating scale (Bond & Fox, 2015). The advantages of the Rasch model in item psychometric development and validation have been documented in the past three decades (Engelhard, 2013; Wright & Masters, 1982; Wright & Stone, 1999). Rasch analysis is invariant as a source of strength. Rasch Model measurement converts ordinal raw data to logit unit data of the same interval data (log odds unit scale) (Bond & Fox, 2015).

Data that is compatible with the Rasch model in this pilot study makes an objective measure of the respondent's ability, and item difficulty can be obtained independently of the item and the respondent. Therefore, this study allows the investigation to be carried out separately between the respondent and the item (Bond & Fox, 2015). Rasch analysis was also conducted to examine the use of rating scales and the displacement of item difficulty according to individual ability.

According to the recommendations of Lim, Rodger, and Brown (2009), every good study should test the psychometric properties of the inventory by implementing the seven main assumptions of Rasch assessment for validity based on test content, response process, and internal structure as outlined in the book Rasch Measurement: Application in Quantitative Educational Research i.e. unidimensionality, measurement scale functionality, item properties, person response, item targeting, reliability and item bias (Myint Swe Khine 2020).

RESULTS

Pilot Study
In the pilot study conducted, a total of 44 science teachers were given the ISTRI-CBA questionnaire with five measurement dimensions consisting of knowledge of CBA, skills of CBA, resource support, attitudes, and professionalism values. All respondents gave good responses and were retained in the study to determine the psychometric characteristics of the items for each dimension in meeting the main assumptions of the Rasch Model.

Dimensionality
The principal component analysis of Rasch residuals is usually used to study the dimensionality of scale. According to Linacre (2019a), the scale will be considered unidimensional if the first dimension explains at least 40% of the variance and the secondary dimension from the unexplained variance of the first contrast is less than 15.0% or has an eigenvalue less than or equal to 2.0, and this indicates that the dimension the largest secondary has a strength of less than two items in one dimension. Eigenvalue comes from the table result analysis by Rasch Model at Unexplained variance in first contrast. Based on the values given, testing was conducted on the five dimensions of the ISTRI-CBA Inventory.

The results for the dimension of knowledge of CBA for ISTRI-CBA show that the variance explained by the measure is 44.4%, and the secondary dimension accounts for 8.7% of the unexplained variance (eigenvalue 7.5). These results provide evidence that the dimension of knowledge of CBA is unidimensional in ISTRI-CBA.

The results for the skill dimension of CBA for ISTRI-CBA show that the variance explained by the measurement is 47.9%, and the secondary dimension accounts for 5.9% of the unexplained variance (eigenvalue 5.1). These results provide evidence that the skill dimension of CBA is unidimensional in ISTRI-CBA.

The results for the resource support dimension for the ISTRI-CBA show that the variance explained by the measure is 46.4%, and the secondary dimension accounts for 9.1% of the unexplained variance (eigenvalue 6.6). These results provide evidence that the resource support dimension is unidimensional in the ISTRI-CBA.

The results for the attitude dimension in PBD for ISTRI-CBA show that the variance explained by the measure is 54.3%, and the secondary dimension accounts for 6.6% of the unexplained variance (eigenvalue 3.7). These results provide evidence that the attitude dimension is unidimensional in ISTRI-CBA.

The results for the professionalism value dimension for ISTRI-CBA show that the variance explained by the measure is 50.1%, and the secondary dimension accounts for 13.8% of the unexplained variance (eigenvalue 2.7). These results provide evidence that the professionalism value dimension is unidimensional in ISTRI-CBA.

Rating Scale Functioning

The effectiveness of the rating scale is to ensure that accurate and stable estimates can be made in the measurement, which is where each category of the rating scale requires at least ten observations (Linacre, 2002). Next, the value of Outfit MnSq should be less than two (Linacre, 2019a). The average value should increase monotonically with the category scale stating that the higher the category scale indicates more latent variables, the measurement scale step for the Andrich threshold should increase monotonically between 1.4 and 5 logit to indicate that people with more latent variables have a higher probability of choosing higher categories on the scale (Linacre, 2002). The improvement of the quality of the use of the measurement scale can also be observed visually on the probability curve, where each category of the scale should show a different peak (Linacre, 2002).

As presented in Table 2 below for the scale functioning rating for the dimension of CBA knowledge, there is two category that has zero frequency count (category one), and categories three to four are observed to have readings of 63, 1087, and 770 frequency counts. The value of Outfit MnSq is in the acceptable range between 0.95 and 1.06. The average measure value increased monotonically from -1.40 to 3.88 across the three categories, and the subsequent determination was by looking at the Andrich threshold reading with a reading between -3.16 and 3.16 logit for categories 3 and 4. The Probability Curve in Figure 3 also shows that the peak of Category One is not visible, while the peaks of Categories two, three, and four are clearly distinct.

Table 2. *Rating Scale Functioning for Knowledge of CBA.*

Category	Frequency count	Average Measure	Outfit MnSq	Andrich threshold
1	0	0	0	0
2	63	-1.40	0.95	None
3	1087	1.10	1.06	-3.16
4	770	3.88	0.96	3.16

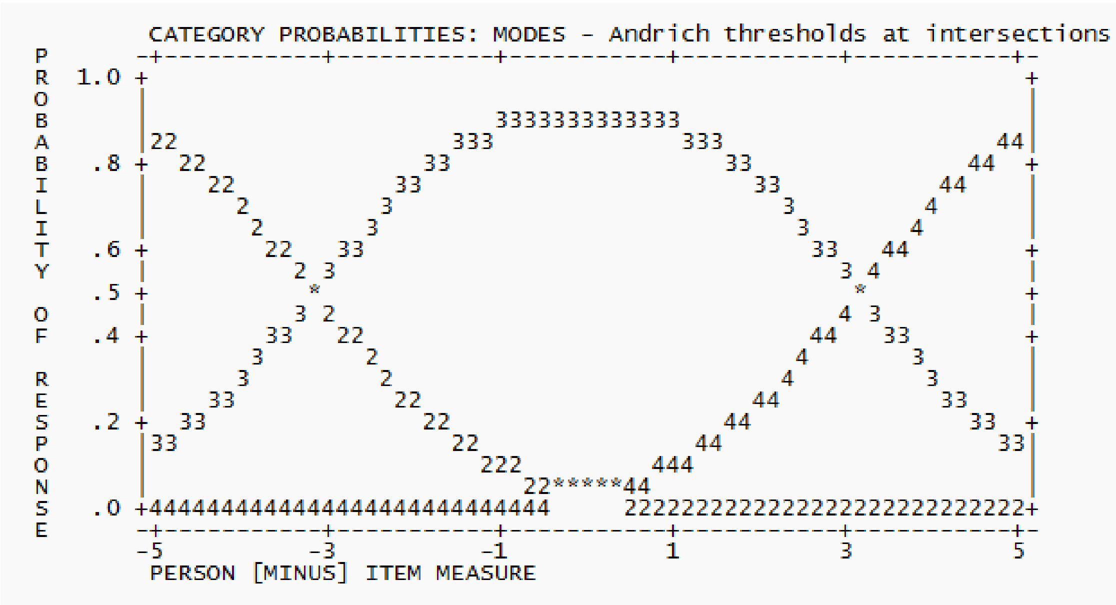

Figure 3. *Probability curves for Knowledge of CBA Scale*

As shown in Table 3 below for the rating scale functioning for the skill dimension of CBA, there is one category that has zero frequency count (category one), and categories two to four are observed to have readings of 37, 1016, and 747 frequency counts. The value of Outfit MnSq is in the acceptable range between 0.62 and 1.01. The average measure reading value increased monotonically from -2.06 to 4.97 across the three categories, and the subsequent determination was by looking at the Andrich threshold reading with readings between -4.09 and 4.09 logit for categories 3 and 4. The Probability Curve in Figure 4 also shows that the peak of Category One is not visible, while the peaks of Categories Two, Three, and Four are clearly distinct.

Table 3. *Rating scale functioning for Skills of CBA*

Category	Frequency count	Average Measure	Outfit MnSq	Andrich threshold
1	0	0	0	0
2	37	-2.06	0.62	None
3	1016	1.53	1.01	-4.09
4	747	4.97	0.80	4.09

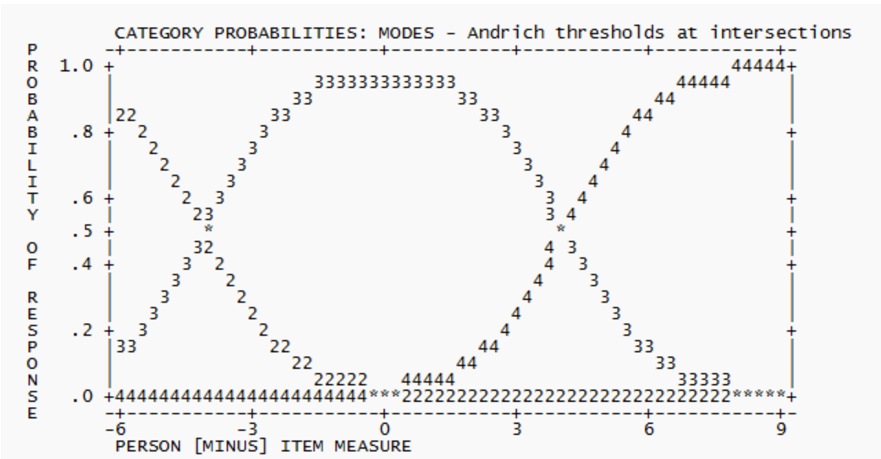

Figure 4. *Probability curves for Skills of CBA Scale*

As shown in Table 4 below for the scale functioning rating for the research support dimension, all scale categories have a frequency of more than 10, namely categories one to four with 19, 265, 839, and 437 frequency counts. The Outfit MnSq value is in an acceptable range between 0.88 and 1.31. The average measure reading value increased monotonically from -0.73 to 3.59 across all four categories, and the subsequent determination was by looking at the Andrich threshold reading with a value between 2.97 and 3.51 logit. The Probability Curve in Figure 5 shows that all the category peaks look clearly different.

Table 4. *Rating scale functioning for Resources Support*

Category	Frequency count	Average Measure	Outfit MnSq	Andrich threshold
1	19	-0.73	1.31	None
2	265	0.06	0.88	-3.15
3	839	1.82	1.04	-0.18
4	437	3.59	1.03	3.33

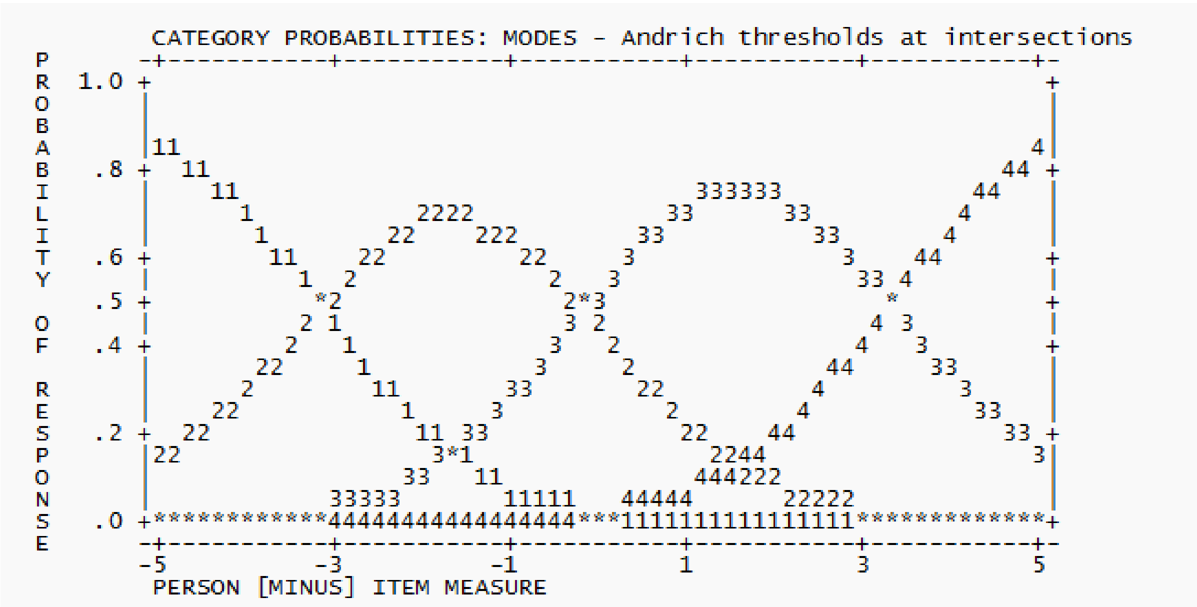

Figure 5. *Probability curves for Resources Support Scale*

As shown in Table 5 below for the scale functioning rating for the attitude dimension, all scale categories have a frequency of more than 10, namely categories one to four with 11, 106, 556, and 445 frequency counts. The Outfit MnSq value is in an acceptable range between 0.67 and 4.66. The average measure reading value increased monotonically from -0.26 to 5.19 across the four categories, and the following determination was by looking at the Andrich threshold reading with a value between 2.84 and 3.38 logit. The Probability Curve in Figure 6 shows that all the category peaks look clearly different.

Table 5. *Rating scale functioning for Attitude*

Category	Frequency count	Average Measure	Outfit MnSq	Andrich threshold
1	11	0.76	4.66	None
2	106	-0.26	0.71	-3.39
3	556	2.11	0.67	-0.55
4	445	5.19	1.04	3.93

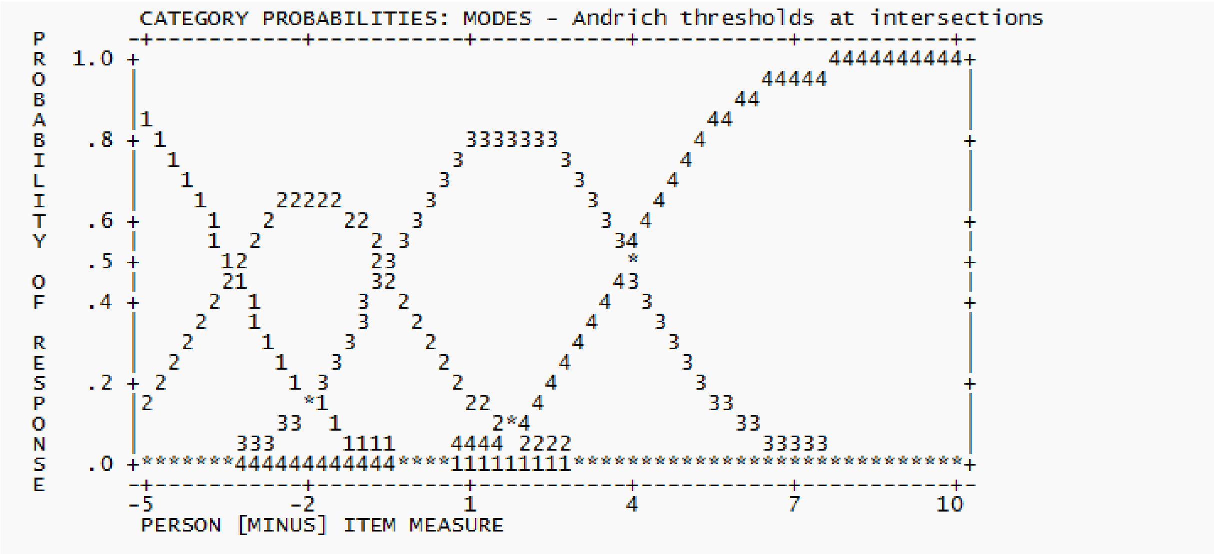

Figure 6. *Probability curves for Attitude Scale*

As shown in Table 6 below for the rating scale functioning for the professionalism value dimension, there is one category that has zero frequency, and categories two to four are observed to have readings of 28, 198, and 104 frequency counts. The value of Outfit MnSq is in the acceptable range between 0.84 and 1.25. The average measure reading value increased monotonically from -1.98 to 3.50 across the three categories, and the subsequent determination was by looking at the Andrich threshold reading with a reading between -2.82 and 5.64 logit. The Probability Curve in Figure 7 also shows that the peak of Category 1 is not visible, while the peaks of Category 2, 3, and 4 are clearly distinct.

Table 6. *Rating scale functioning for professionalism value*

Category	Frequency count	Average Measure	Outfit MnSq	Andrich threshold
1	0	0	0	0
2	28	-1.98	1.25	None
3	198	0.60	0.98	-2.82
4	104	3.50	0.84	2.82

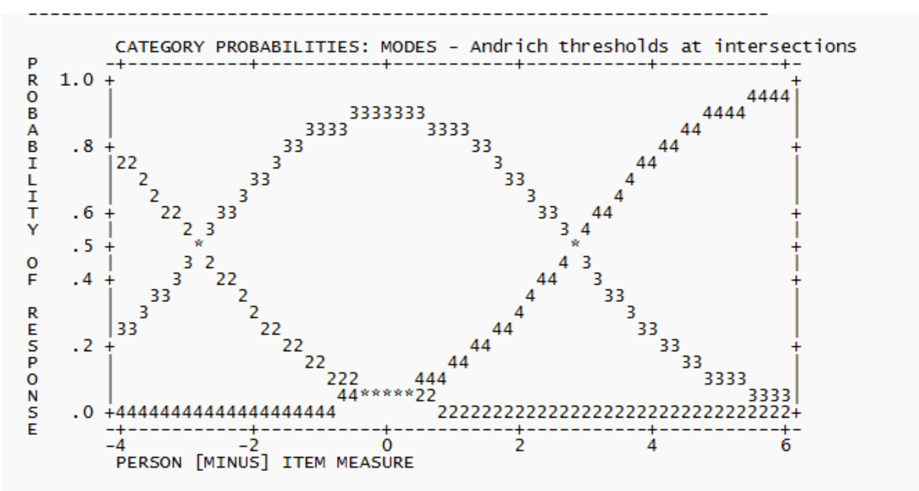

Figure 7. *Probability curves for professionalism value Scale*

Item Properties

Item quality can be checked through several indicators. Local Independence of items is an assumption in the Rasch model, meaning that the items in a test should not be related to each other (Baghaei, 2008). Standardized residual correlations reflect local dependence between items, where a correlation greater than 0.7 suggests that two items share more than half of their random variance, and only one should be retained (Linacre, 2019a). Item measurement correlation refers to the correlation between observations and Rasch measurements; it serves as a check that the respondent's scoring makes sense. Note that the item measure correlation, which is expected to be positive, is used to determine whether the response level scoring for each item is the same orientation as the latent variable (Linacre, 2019a). Negative correlations suggest reverse-coded item errors, while near-zero correlations suggest items are too easy or challenging to measure in different dimensions. A correlation of less than 0.4 can be used to flag items related to the use of instrument words (Wolfe & Smith, 2007).

Item fit statistics, including Outfit and Infit MnSq and t-statistics (Zstd), estimate how well empirical data fit the Rasch model (Bond & Fox, 2015). Although the best MnSq is 1, values between 0.5 and 1.5 (John, 1994) with Zstd between −2 and 2 are accepted as a reasonable fit (Bond & Fox, 2015).

Findings for the knowledge dimension of CBA in item quality, for the most significant standardized residual correlations, have a satisfactory range from 0.53 to 0.81 (one less satisfactory item), respectively. Table 7 shows the average reading of the Outfit MnSq item value is at the ideal value of 1.0 for four items, good items for the Outfit MnSq reading range from 0.5 to 1.5 (37 items), items that can be improved with Zstd reading values between −2 and 2 (four items). In comparison, about three items outside the reading of Outfit MnSq and Zstd were considered to be dropped.

Findings for the skill dimension of CBA in item quality, for the most significant standardized residual correlations, have a satisfactory range from 0.48 to 0.93 (one less satisfactory item). Table 8 indicates that the average reading of the value of the Outfit MnSq item is at the ideal value of 1.0 for two items, a good item for the reading of Outfit MnSq ranges from 0.5 to 1.5 (36 items), an item that can be repaired with a Zstd reading value between −2 and 2 (six items) while one item outside the reading of Outfit MnSq and Zstd was considered to be dropped.

Findings for the resource support dimension in item quality, for the most extensive standardized residual correlations, have a satisfactory range from 0.52 to 0.80 (one less satisfactory item). Table 9 indicates that the average reading of the value of the Outfit MnSq item is at the ideal value of 1.0 for about eight items, for items that are good for reading Outfit MnSq ranging from 0.5 to 1.5 for 29 items, while about two items outside the reading of Outfit MnSq and Zstd were considered to be aborted.

Findings for the attitude dimension in item quality, for the most extensive standardized residual correlations of all items, have a satisfactory range from 0.39 to 0.68. Table 10 shows that there is no reading of the average value of the Outfit MnSq item at the ideal value of 1.0. as about 16 good items for Outfit MnSq readings ranging from 0.5 to 1.5, items can be repaired with Zstd readings between −2 and 2, which are six items while four items that are outside the Outfit MnSq and Zstd readings were considered to be dropped.

Findings for the attitude dimension in item quality, for the most significant standardized residual correlations of all items, have a satisfactory range from 0.27 to 0.61 (one item). Table 11 shows that there is one item with an average reading of the value of the Outfit MnSq item at the ideal value of 1.0. Among seven good items for Outfit MnSq readings ranging from 0.5 to 1.5, one item can be repaired with a Zstd reading value between −2 and 2. In comparison, one item that is outside the Outfit MnSq and Zstd readings was considered to be dropped.

Table 7. *Item measure estimates, measure correlations and fit statistics for Knowledge of CBA*

TABLE 13.1 Miss fit_PENGETAHUAN PBD_40.xlsx ZOU156WS.TXT Sep 21 2022 10:39
INPUT: 40 PERSON 48 ITEM REPORTED: 40 PERSON 48 ITEM 3 CATS WINSTEPS 5.1.0.0

PERSON: REAL SEP.: 4.06 REL.: .94 ... ITEM: REAL SEP.: 1.72 REL.: .75

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
32	122	40	2.10	.41	1.89	2.96	2.03	2.46	.55	.71	59.5	78.2	I32
17	124	40	1.76	.41	1.50	1.89	1.59	1.66	.66	.70	64.9	77.7	I17
19	127	40	1.27	.40	.86	-.55	.94	-.08	.66	.68	83.8	77.2	I19
20	127	40	1.27	.40	.79	-.90	.68	-1.03	.70	.68	78.4	77.2	I20
39	128	40	1.11	.40	1.24	1.06	1.33	1.04	.64	.68	70.3	77.0	I39
11	130	40	.78	.40	1.04	.27	.97	.01	.67	.67	75.7	76.5	I11
12	130	40	.78	.40	1.07	.39	.94	-.07	.72	.67	70.3	76.5	I12
16	131	40	.62	.40	1.18	.85	1.26	.84	.63	.67	73.0	76.1	I16
21	131	40	.62	.40	1.29	1.28	1.14	.52	.65	.67	73.0	76.1	I21
34	131	40	.62	.40	.66	-1.68	.50	-1.79	.78	.67	83.8	76.1	I34
45	131	40	.62	.40	.67	-1.61	.67	-1.06	.66	.67	83.8	76.1	I45
31	132	40	.46	.40	.55	-2.42	.43	-2.12	.72	.67	91.9	75.8	I31
33	132	40	.46	.40	.69	-1.56	.52	-1.68	.73	.67	75.7	75.8	I33
44	132	40	.46	.40	.74	-1.23	.70	-.93	.71	.67	86.5	75.8	I44
46	132	40	.46	.40	.76	-1.12	.75	-.73	.65	.67	86.5	75.8	I46
48	132	40	.46	.40	1.03	.23	1.07	.32	.68	.67	75.7	75.8	I48
4	133	40	.30	.40	1.16	.77	1.23	.76	.60	.66	73.0	75.5	I4
13	133	40	.30	.40	1.43	1.85	1.49	1.39	.65	.66	62.2	75.5	I13
24	133	40	.30	.40	.72	-1.39	.58	-1.40	.73	.66	83.8	75.5	I24
47	133	40	.30	.40	.78	-1.06	.65	-1.11	.71	.66	78.4	75.5	I47
15	134	40	.14	.40	1.32	1.44	1.39	1.16	.66	.66	70.3	75.7	I15
28	134	40	.14	.40	.68	-1.59	.54	-1.56	.75	.66	86.5	75.7	I28
42	134	40	.14	.40	1.02	.16	.90	-.20	.66	.66	75.7	75.7	I42
2	135	40	-.02	.40	.89	-.44	.69	-.92	.71	.66	70.3	76.0	I2
26	135	40	-.02	.40	.71	-1.40	.56	-1.42	.74	.66	81.1	76.0	I26
30	135	40	-.02	.40	.80	-.90	.67	-.99	.72	.66	81.1	76.0	I30
35	135	40	-.02	.40	.71	-1.42	.63	-1.14	.69	.66	86.5	76.0	I35
41	135	40	-.02	.40	.84	-.70	.81	-.47	.65	.66	81.1	76.0	I41
3	136	40	-.19	.41	1.13	.64	1.36	1.05	.57	.66	73.0	76.2	I3
14	136	40	-.19	.41	1.08	.42	1.03	.21	.64	.66	67.6	76.2	I14
18	136	40	-.19	.41	.97	-.06	.86	-.31	.69	.66	73.0	76.2	I18
22	136	40	-.19	.41	1.24	1.08	1.59	1.56	.68	.66	73.0	76.2	I22
27	136	40	-.19	.41	1.02	.18	1.18	.61	.66	.66	73.0	76.2	I27
38	136	40	-.19	.41	1.17	.82	1.25	.79	.62	.66	73.0	76.2	I38
5	137	40	-.35	.41	.88	-.48	.72	-.77	.72	.66	75.7	76.7	I5
40	137	40	-.35	.41	1.14	.68	1.27	.82	.63	.66	75.7	76.7	I40
29	138	40	-.52	.41	.54	-2.34	.40	-2.06	.75	.66	94.6	77.5	I29
37	138	40	-.52	.41	1.06	.35	1.10	.38	.61	.66	78.4	77.5	I37
43	138	40	-.52	.41	1.17	.80	1.31	.91	.57	.66	73.0	77.5	I43
8	139	40	-.69	.42	1.47	1.82	1.45	1.19	.62	.66	62.2	78.1	I8
25	140	40	-.87	.42	.83	-.65	.91	-.13	.70	.66	86.5	78.9	I25
36	140	40	-.87	.42	.93	-.23	.88	-.19	.65	.66	81.1	78.9	I36
23	141	40	-1.05	.43	.77	-.91	.59	-1.08	.70	.65	81.1	80.0	I23
1	142	40	-1.24	.43	.57	-1.87	.48	-1.41	.74	.65	94.6	80.9	I1
6	144	40	-1.62	.45	1.00	.10	1.15	.46	.62	.64	86.5	82.1	I6
9	144	40	-1.62	.45	1.28	1.03	2.01	1.81	.54	.64	75.7	82.1	I9
10	145	40	-1.82	.45	1.46	1.55	1.59	1.14	.57	.63	73.0	82.4	I10
7	147	40	-2.24	.46	1.05	.26	1.04	.28	.59	.62	81.1	82.8	I7
MEAN	134.7	40.0	.00	.41	1.00	-.1	1.00	-.1			77.4	77.1	
P.SD	5.1	.0	.86	.01	.28	1.2	.40	1.1			7.9	2.0	

Table 8. *Item measure estimates, measure correlations and fit statistics for Skills of CBA*

TABLE 13.1 MISS FIT_KEMAHIRAN PBD.xlsx ZOU587WS.TXT Sep 21 2022 17:27
INPUT: 40 PERSON 45 ITEM REPORTED: 40 PERSON 45 ITEM 3 CATS WINSTEPS 5.1.0.0

PERSON: REAL SEP.: 4.32 REL.: .95 ... ITEM: REAL SEP.: 1.63 REL.: .73

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
15	125	40	2.02	.46	.82	-.59	.61	-.92	.75	.69	83.8	82.8	I15
30	125	40	2.02	.46	2.04	2.97	2.54	2.71	.59	.69	67.6	82.8	I30
17	126	40	1.81	.45	.77	-.83	.69	-.72	.71	.68	89.2	82.0	I17
36	126	40	1.81	.45	.99	.06	.72	-.64	.67	.68	78.4	82.0	I36
11	129	40	1.22	.43	1.65	2.38	1.73	1.65	.55	.67	70.3	79.8	I11
18	129	40	1.22	.43	.92	-.29	.82	-.35	.65	.67	81.1	79.8	I18
9	132	40	.67	.42	1.38	1.68	1.30	.77	.67	.67	75.7	77.7	I9
19	132	40	.67	.42	1.08	.43	.88	-.14	.65	.67	81.1	77.7	I19
38	132	40	.67	.42	.79	-1.00	.58	-.99	.65	.67	81.1	77.7	I38
16	133	40	.50	.42	1.20	.97	.93	.00	.64	.67	62.2	77.3	I16
37	133	40	.50	.42	.86	-.68	.68	-.63	.64	.67	78.4	77.3	I37
39	133	40	.50	.42	.66	-1.83	.46	-1.36	.69	.67	89.2	77.3	I39
40	133	40	.50	.42	.88	-.55	.70	-.58	.63	.67	83.8	77.3	I40
29	134	40	.32	.42	1.06	.34	.85	-.17	.61	.67	75.7	77.3	I29
41	134	40	.32	.42	.92	-.36	.81	-.29	.76	.67	81.1	77.3	I41
42	134	40	.32	.42	.85	-.72	.65	-.70	.72	.67	81.1	77.3	I42
43	134	40	.32	.42	1.00	.08	.94	.02	.68	.67	75.7	77.3	I43
24	136	40	-.03	.42	.80	-1.05	.55	-.94	.75	.67	83.8	77.6	I24
25	136	40	-.03	.42	1.06	.39	1.50	1.05	.61	.67	78.4	77.6	I25
28	136	40	-.03	.42	1.10	.55	.88	-.09	.62	.67	73.0	77.6	I28
32	136	40	-.03	.42	.80	-1.04	.55	-.95	.75	.67	78.4	77.6	I32
4	137	40	-.21	.42	1.14	.74	1.14	.44	.68	.68	75.7	77.6	I4
5	137	40	-.21	.42	1.04	.27	.86	-.12	.70	.68	75.7	77.6	I5
12	137	40	-.21	.42	.99	.01	.83	-.19	.71	.68	81.1	77.6	I12
14	137	40	-.21	.42	.65	-1.95	.45	-1.25	.78	.68	86.5	77.6	I14
22	137	40	-.21	.42	1.02	.16	.89	-.05	.64	.68	81.1	77.6	I22
23	137	40	-.21	.42	.78	-1.18	.66	-.62	.69	.68	86.5	77.6	I23
33	137	40	-.21	.42	.77	-1.19	.52	-1.01	.70	.68	81.1	77.6	I33
45	137	40	-.21	.42	.94	-.25	.63	-.69	.67	.68	86.5	77.6	I45
26	138	40	-.39	.43	.87	-.61	.63	-.69	.69	.68	78.4	77.9	I26
8	139	40	-.58	.43	1.02	.18	.73	-.42	.67	.68	78.4	78.1	I8
13	139	40	-.58	.43	.70	-1.58	.47	-1.15	.79	.68	89.2	78.1	I13
20	139	40	-.58	.43	1.19	.94	1.32	.74	.68	.68	73.0	78.1	I20
35	139	40	-.58	.43	1.21	1.03	.79	-.28	.64	.68	78.4	78.1	I35
44	139	40	-.58	.43	.81	-.93	.54	-.94	.77	.68	78.4	78.1	I44
2	140	40	-.77	.44	1.02	.18	.74	-.38	.67	.69	75.7	79.0	I2
3	141	40	-.97	.45	.99	.02	.71	-.44	.69	.69	78.4	80.3	I3
10	141	40	-.97	.45	.69	-1.42	.46	-1.13	.75	.69	89.2	80.3	I10
21	141	40	-.97	.45	1.00	.06	.99	.16	.67	.69	83.8	80.3	I21
27	141	40	-.97	.45	.87	-.51	.56	-.82	.72	.69	78.4	80.3	I27
31	141	40	-.97	.45	.85	-.62	.58	-.78	.72	.69	83.8	80.3	I31
34	141	40	-.97	.45	.82	-.75	.56	-.83	.72	.69	83.8	80.3	I34
1	142	40	-1.17	.46	1.46	1.68	2.22	1.87	.56	.70	75.7	81.7	I1
7	142	40	-1.17	.46	.85	-.53	.71	-.43	.72	.70	86.5	81.7	I7
6	143	40	-1.39	.48	.84	-.51	.77	-.25	.72	.70	89.2	83.4	I6
MEAN	135.8	40.0	.00	.43	.98	-.1	.85	-.3			80.1	78.8	
P.SD	4.7	.0	.86	.01	.26	1.0	.43	.8			5.8	1.8	

Table 9. *Item measure estimates, measure correlations and fit statistics for Resources Support*

TABLE 13.1 SOKONGAN SUMBER_3.xlsx ZOU883WS.TXT Nov 26 2022 17:23
INPUT: 40 PERSON 39 ITEM REPORTED: 40 PERSON 39 ITEM 4 CATS WINSTEPS 5.1.0.0

PERSON: REAL SEP.: 3.87 REL.: .94 ... ITEM: REAL SEP.: 3.31 REL.: .92

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
4	93	40	2.63	.28	1.25	1.11	1.23	1.03	.52	.61	57.5	64.5	I4
31	95	40	2.46	.28	1.88	3.22	1.94	3.38	.35	.61	47.5	64.7	I31
39	99	40	2.14	.29	1.09	.45	1.09	.45	.52	.61	62.5	64.3	I39
29	103	40	1.81	.29	1.21	.96	1.26	1.16	.56	.60	52.5	63.2	I29
30	107	40	1.48	.29	1.41	1.75	1.49	2.01	.66	.60	60.0	63.0	I30
35	109	40	1.31	.29	.64	-1.87	.66	-1.72	.67	.59	85.0	63.4	I35
5	112	40	1.06	.29	.60	-2.14	.59	-2.18	.77	.59	75.0	63.1	I5
36	112	40	1.06	.29	1.35	1.53	1.40	1.73	.64	.59	62.5	63.1	I36
37	112	40	1.06	.29	.69	-1.56	.71	-1.43	.65	.59	70.0	63.1	I37
10	116	40	.72	.29	1.02	.17	1.02	.18	.44	.58	55.0	62.7	I10
3	118	40	.55	.30	.96	-.12	.97	-.08	.63	.58	50.0	63.4	I3
9	120	40	.37	.30	1.02	.16	.98	.00	.61	.58	62.5	64.0	I9
2	121	40	.28	.30	.74	-1.27	.73	-1.27	.62	.58	80.0	64.5	I2
8	122	40	.19	.30	1.13	.63	1.14	.67	.55	.58	65.0	64.9	I8
33	122	40	.19	.30	1.23	1.04	1.18	.83	.74	.58	45.0	64.9	I33
1	125	40	-.09	.31	1.54	2.14	1.50	1.98	.58	.58	57.5	66.6	I1
6	125	40	-.09	.31	1.14	.69	1.09	.46	.65	.58	62.5	66.6	I6
32	125	40	-.09	.31	1.09	.46	1.04	.26	.61	.58	62.5	66.6	I32
7	126	40	-.18	.31	1.03	.23	.99	.05	.66	.58	72.5	67.2	I7
18	126	40	-.18	.31	.67	-1.60	.66	-1.61	.68	.58	77.5	67.2	I18
25	128	40	-.37	.31	.93	-.26	.92	-.26	.42	.57	67.5	68.5	I25
34	128	40	-.37	.31	1.20	.88	1.16	.73	.75	.57	57.5	68.5	I34
16	129	40	-.47	.31	.62	-1.83	.63	-1.69	.58	.57	75.0	68.9	I16
38	129	40	-.47	.31	1.05	.28	1.20	.87	.59	.57	60.0	68.9	I38
15	130	40	-.57	.31	1.01	.11	1.21	.89	.30	.57	67.5	69.2	I15
17	130	40	-.57	.31	.48	-2.77	.48	-2.57	.70	.57	82.5	69.2	I17
23	130	40	-.57	.31	.70	-1.38	.67	-1.46	.69	.57	72.5	69.2	I23
27	130	40	-.57	.31	.84	-.67	.82	-.70	.68	.57	72.5	69.2	I27
24	132	40	-.77	.32	.86	-.55	.86	-.52	.56	.57	67.5	69.8	I24
28	132	40	-.77	.32	1.15	.68	1.10	.48	.55	.57	67.5	69.8	I28
13	133	40	-.87	.32	.65	-1.63	.94	-.14	.61	.56	82.5	70.2	I13
14	133	40	-.87	.32	.77	-1.03	1.09	.43	.44	.56	77.5	70.2	I14
26	133	40	-.87	.32	.76	-1.04	.72	-1.12	.65	.56	82.5	70.2	I26
22	134	40	-.97	.32	.94	-.19	.94	-.14	.48	.56	72.5	70.5	I22
21	135	40	-1.08	.32	.99	.05	.99	.06	.54	.56	70.0	70.6	I21
12	137	40	-1.29	.33	.70	-1.42	.64	-1.39	.60	.55	82.5	70.8	I12
20	137	40	-1.29	.33	1.03	.22	1.02	.16	.38	.55	62.5	70.8	I20
11	139	40	-1.51	.33	1.02	.15	1.03	.18	.41	.53	72.5	70.4	I11
19	147	40	-2.48	.37	1.06	.34	.96	.06	.39	.45	62.5	73.6	I19
MEAN	123.4	40.0	.00	.31	.99	-.1	1.00	.0			67.1	67.2	
P.SD	12.3	.0	1.12	.02	.28	1.3	.29	1.2			10.1	3.0	

Table 10. *Item measure estimates, measure correlations and fit statistics for Attitude*

TABLE 13.1 SIKAP.xlsx ZOU996WS.TXT Sep 23 2022 15:33
INPUT: 44 PERSON 26 ITEM REPORTED: 44 PERSON 26 ITEM 4 CATS WINSTEPS 5.1.0.0

PERSON: REAL SEP.: 3.31 REL.: .92 ... ITEM: REAL SEP.: 2.60 REL.: .87

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
26	102	44	3.97	.29	6.44	9.91	7.79	9.91	-.12	.79	31.0	67.8	I26
10	136	44	.98	.32	.48	-2.71	.45	-2.78	.83	.72	88.1	71.3	I10
15	137	44	.88	.32	.73	-1.18	.74	-1.06	.78	.72	76.2	71.9	I15
25	137	44	.88	.32	.61	-1.85	.62	-1.71	.84	.72	85.7	71.9	I25
11	139	44	.67	.33	.95	-.15	.99	.03	.75	.71	81.0	73.1	I11
7	140	44	.56	.33	.86	-.54	.77	-.84	.74	.71	78.6	73.4	I7
13	142	44	.33	.34	.54	-2.20	.52	-2.05	.79	.71	83.3	73.7	I13
22	142	44	.33	.34	.65	-1.57	.60	-1.59	.82	.71	83.3	73.7	I22
23	142	44	.33	.34	.40	-3.11	.35	-3.10	.85	.71	92.9	73.7	I23
6	143	44	.22	.34	.66	-1.47	.66	-1.25	.79	.70	90.5	73.7	I6
24	143	44	.22	.34	.72	-1.15	.66	-1.28	.80	.70	81.0	73.7	I24
12	144	44	.10	.34	.88	-.43	.79	-.69	.77	.70	73.8	74.3	I12
14	144	44	.10	.34	.46	-2.66	.47	-2.20	.80	.70	92.9	74.3	I14
3	145	44	-.02	.35	.54	-2.16	.47	-2.13	.83	.70	88.1	74.8	I3
5	145	44	-.02	.35	.85	-.53	.76	-.78	.79	.70	69.0	74.8	I5
17	147	44	-.26	.35	.79	-.81	.71	-.93	.74	.69	76.2	75.8	I17
8	149	44	-.51	.35	.62	-1.72	.51	-1.66	.79	.68	85.7	76.2	I8
16	149	44	-.51	.35	.79	-.85	.73	-.77	.70	.68	81.0	76.2	I16
4	151	44	-.76	.36	2.75	5.01	2.53	3.06	.22	.66	50.0	76.5	I4
9	151	44	-.76	.36	.81	-.76	.71	-.75	.74	.66	76.2	76.5	I9
20	151	44	-.76	.36	.56	-2.07	.47	-1.66	.77	.66	90.5	76.5	I20
19	153	44	-1.02	.36	.67	-1.51	.54	-1.22	.75	.65	85.7	76.2	I19
21	153	44	-1.02	.36	.49	-2.60	.38	-1.88	.76	.65	85.7	76.2	I21
18	154	44	-1.15	.36	.50	-2.56	.39	-1.72	.75	.64	88.1	75.8	I18
2	155	44	-1.28	.37	.54	-2.36	.41	-1.53	.75	.63	90.5	75.6	I2
1	157	44	-1.55	.37	.70	-1.44	.54	-.89	.67	.61	81.0	75.8	I1
MEAN	144.3	44.0	.00	.34	.96	-.9	.95	-.8			80.2	74.4	
P.SD	10.3	.0	1.06	.02	1.18	2.6	1.43	2.4			13.2	2.0	

Table 11. *Item measure estimates, measure correlations and fit statistics for Professionalism Value*

TABLE 13.1 NILAI PROFESIONALISME_4.xlsx ZOU969WS.TXT Sep 29 2022 17:47
INPUT: 33 PERSON 10 ITEM REPORTED: 33 PERSON 10 ITEM 3 CATS WINSTEPS 5.1.0.0

PERSON: REAL SEP.: 2.35 REL.: .85 ... ITEM: REAL SEP.: 1.93 REL.: .79

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
3	91	33	2.66	.43	2.79	4.85	3.31	4.14	.21	.66	51.5	76.7	I3
1	104	33	.44	.41	1.04	.24	1.10	.42	.70	.65	78.8	74.0	I1
4	106	33	.11	.41	.93	-.25	.86	-.35	.86	.66	75.8	74.3	I4
5	107	33	-.06	.41	.62	-1.91	.49	-1.88	.85	.66	84.8	74.3	I5
8	107	33	-.06	.41	.81	-.83	.66	-1.12	.56	.66	72.7	74.3	I8
6	108	33	-.23	.41	.77	-1.02	.63	-1.20	.68	.66	75.8	74.5	I6
9	108	33	-.23	.41	1.03	.20	.88	-.28	.59	.66	69.7	74.5	I9
7	110	33	-.57	.42	.77	-.98	.65	-1.04	.72	.67	81.8	74.9	I7
2	112	33	-.93	.43	.73	-1.09	.65	-1.00	.76	.68	81.8	77.0	I2
10	113	33	-1.12	.43	.48	-2.44	.36	-2.17	.79	.68	90.9	78.2	I10
MEAN	106.6	33.0	.00	.42	1.00	-.3	.96	-.4			76.4	75.3	
P.SD	5.8	.0	.99	.01	.62	1.9	.81	1.7			10.1	1.4	

Targeting

Item-person targeting investigates the extent to which the distribution of item difficulty matches the range of abilities of an individual (Linacre, 2019a). Findings that show the percentage of respondents who get the lowest score (floor effect) or the highest (ceiling effect) indicate that the items are too easy or difficult for the study respondents. Floor and ceiling effects below 1% are excellent, and effects between 1% and 2% are sound (Fisher, 2007).

A comparison of the mean and standard deviation (SD) of item difficulty and respondent ability should also match to show good displacement (Bond & Fox, 2015). Such evidence can be observed visually in the person wright-map, which traces individual abilities on the left and item difficulties on the right along latent variables. A well-targeted item scale is one in which the difficulty of the items on the Item column covers the entire range of the respondent's abilities (Linacre, 2019b).

In comparison of ISTRI-CBA findings for the item, the mean measurement is 0.07 and SD of 0.83, while the mean measurement of respondents is 0.25 with SD 1.64. The mean item size is lower, and the smaller SD shows the effect on the Wright-map person, where the item difficulty only ranges between the mean and SD on the respondent's ability (Figure 8). Findings in the Wright Map in Figure 8 indicates that the maximum value of the item is lower at the level of +3.50 logit compared to the maximum of a person, which is +7.0 logit. For the lower level, it is found that the item minimum is lower at the -1.92-logit level compared to the person minimum at the +0.44 level. Therefore, it was found that there is no complicated item that measures Person 1. The item only measures Person 2 and Person 3. The right side of this map shows the level of difficulty for each item, from elementary level items (logit value -1.92 for SOK19 item) to difficult level items (logit value 3.50 for SOK4 item). The left side of the map shows the distribution of the logit value of science teacher respondents from a low ability level (person logit value <2.03) to a high ability level (person logit value >3.5). The conclusion was that the items were divided into three categories, namely very easy, within the target, and moderately challenging, based on the ability of the science teacher (person) to answer in the item-individual map. Therefore, each of these categories in the item-individual mapping analysis can check the quality of the items produced in separating individuals according to the difficulty level of the item and the ability of the respondent. Based on the findings in the item-individual map, most people are in the straightforward ability category, within the target, and moderately complex.

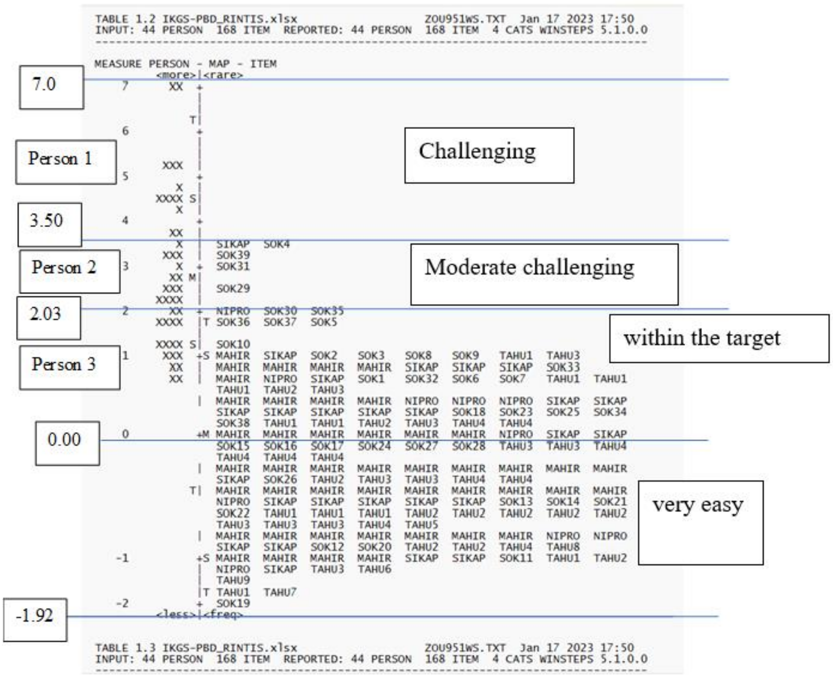

Figure 8. *Item-Person Wright-Map ISTRI-CBA*

Reliability

The reliability index shows the repeatability of person order if the same sample completes a set of items and similar items by showing the reliability of the repeatability of item order if the set of items is assigned to the same sample (Bond & Fox, 2015). Person and item strata indices represent the number of statistically different person levels of ability and item difficulty, respectively, and are calculated as (4 G + 1)/3, where G refers to the person or item separation index (Wright & Masters, 1982). A low-person separation (strata) suggests that the scale is not sensitive enough to distinguish different levels of a person's ability. In contrast, a low item separation (strata) suggests the sample is not large enough to confirm the hierarchy of item difficulty (Linacre, 2019a). Fisher (2007) recommends reliability and strata of at least 0.67 and 2 for fair quality and at least 0.81 and 3 for good quality.

The research findings for the knowledge dimension of CBA have an item reliability of 0.75 with a separation index of 1.72 while the person reliability is 0.96 with a separation index of 5.12. This finding shows that the knowledge dimension of CBA has item reliability and a separation index of fair quality, person reliability, and an excellent separation index.

The findings of the study for the skill dimension of CBA have item reliability of 0.73 with a separation index of 1.63, while person reliability is 0.96 with a separation index of 5.16. This finding shows that the skill dimension of CBA has item reliability and a separation index of fair quality, person reliability, and a perfect separation index.

The research findings for the resource support dimension have item reliability of 0.92 with a separation index of 3.81, while person reliability is 0.94 with a separation index of 3.87. This finding shows that the source support dimension has item reliability and separation index of excellent quality, as well as person reliability and separation index of outstanding quality.

The findings of the study for the attitude dimension have item reliability of 0.87 with a separation index of 2.58, while person reliability is 0.93 with a separation index of 3.54. This finding shows that the attitude dimension has item reliability and separation index of excellent quality, as well as person reliability and separation index of outstanding quality.

Research findings for the value dimension of professionalism have item reliability of 0.79 with a separation index of 1.93 while person reliability is 0.86 with a separation index of 2.35. This finding shows that the value dimension of professionalism has item reliability and a separation index of fair quality, person reliability, and a perfect separation index.

DISCUSSION

The study on the willingness of science teachers to implement CBA is new in Malaysia, where there is limited evidence about the difficulty of obtaining a valid inventory in Malaysia for the purpose of measuring the willingness of science teachers to implement CBA. The CBA that began to be implemented in primary schools in Malaysia in 2019 requires the willingness of teachers to make it a success (Ministry of Education Malaysia, 2018a). Moreover, no research evidence is best developed and validated based on Rasch model scale validation methods. Therefore, this study aims to validate the ISTRI-CBA for science subjects through Rasch assessment. Results from the Study prove that the consistent validity of the ISTRI-CBA is based on test content, response process, and internal structure. Findings suggest that the scale is a valid tool to measure the willingness of science teachers to implement CBA.

Principle component analysis of residuals shows that all dimensions in ISTRI-CBA are unidimensional, with the primary dimension explaining almost 50% of the variation in persons' responses. These results show all dimensions in that the items do not create secondary factors in the ISTRI-CBA (Wong, Rindfleisch, & Burroughs, 2003). Observations on the use of rating scales suggest that the 4-point scale used is less effective on a category-one scale. However, all categories were used sufficiently to provide accurate and stable estimates, and the observed noise did not degrade the measurements. The monotonic increase in measurement and calibration of steps up the rating scale also provides evidence for the rating scale theory that higher categories indicate more latent variables and that people with more latent variables have a higher chance of choosing higher categories. Various aspects of the item's properties confirm that the item works as expected. Satisfactory correlations of residual and standardized item measures suggest that all items are independent of each other (Linacre, 2019a).

The excellent item reliability and strata-separated index indicate that the sample is sufficient to confirm that the observed item order difficulties are likely to be reproduced in other similar samples. This statement supports that items with high step estimates are indeed more complex than items with lower step estimates. Consequently, the fair response reliability index suggests a lower probability of observing the same person's ability to order if the same sample is given the same measure. The observed separation strata person index is more than three, which shows that the scale can distinguish three levels of a person's abilities (Smith et al., 2013). This result shows the displacement findings above are recommending items that are sufficient for better reliability for a person with a different level of personal ability.

Limitations and Future Research

The current study significantly contributes to the applicability of the ISTRI-CBA rating scale with multiple, larger samples. However, since our sample is limited to only 44 science teachers, further validation with a more diverse sample of schools in Malaysia or other countries is required in order to draw more confident conclusions about the generalizability of ISTRI-CBA to educators in Malaysia.

Results from person-item targeting, as well as person-reliability and strata-separated indices, suggest that the ISTRI-CBA could benefit from additional items at higher levels of difficulty (Smith et al., 2008), Future improvements of the ISTRI-CBA should explore more items to target students with higher abilities and increase ability level differences. Item displacement can also be made with fewer items by focusing only one item on each respondent's ability. This finding is to ensure a reasonable length of the scale. At the same time, items with the same level of difficulty can be revised to increase difficulty based on theory and can be rearranged for a better spread of difficulty.

Another angle of future research that needs to be considered is the use of individual readiness theory to understand further whether the readiness of science teachers to implement CBA in schools is present in every science teacher in order to implement change. Identifying the level of readiness of science teachers can provide understanding on how it can help the development of interventions in CBA for science teachers in schools.

CONCLUSION

This study is the pilot study to evaluate ISTRI-CBA through Rasch analysis with the involvement of a small sample size. Findings across studies provide empirical evidence that consistently supports the ISTRI-CBA as valid and sufficiently reliable instruments to measure teachers' readiness to implement CBA in science subjects. Future improvements of the ISTRI-CBA may consider adding or reducing items at higher difficulty levels to better target different ability levels and with a bigger sample size.

REFERENCES

Abdul Said Ambotang, & Shanti Gobalakrishnan. (2017). Tahap integriti dan kesediaan guru dengan pelaksanaan Pentaksiran Berasaskan Sekolah Rendah. *International Journal of Education, Psychology and Counseling*, *Volume 2*(4), pages 1-22.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43(4), 561–573.*

Arar, K., & Abramovitz, R. (2017). Teacher-related factors in assimilation of technological change in schools. *International Journal of Educational Management*.

Arsaythamby Veloo, Hariharan N. Krishnasamy, & Ruzlan Md-Ali. (2015). Teachers' Knowledge and Readiness towards Implementation of School Based Assessment in Secondary Schools.

International Education Studies, *8*(11), 193-203. https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1082112&site=eds-live.

Avidov-Ungar, O., & Arviv-Elyashiv, R. (2020). Teachers' perceptions of educational reform: the schools' readiness, supporting mechanisms and contributions of the reform. *International Journal of Educational Management*.

Baghaei, P. (2008). Local dependency and Rasch measures. *Rasch Measurement Transactions, 21(3)*, 1105–1106. Retrieved from https://www.rasch.org/rmt/rmt213b.htm

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.).* New York, NY: Routledge.

Bouillet, D., Ivanec, T. P., & Miljević-Riđički, R. (2014). Preschool teachers' resilience and their readiness for building children's resilience. *Health Education*.

Bruce, B. F. (2018b). The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation [Language material]. Sage Publications. https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edscrc&AN=edscrc.26960446&site=eds-live.

Christensen, R., & Knezek, G. (2017). Validating a mobile learning readiness survey: Assessing teachers' dispositions toward adoption. *Journal of Digital Learning in Teacher Education*, *33*(4), 148-159. https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1154582&site=eds-live http://dx.doi.org/10.1080/21532974.2017.1347536.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge.

Fakhri Abdul Khalil, & Mohd Isha Awang. (2016). Isu kesediaan guru dalam amalan melaksanakan pentaksiran berasaskan sekolah. *EDUCATUM Journal of Social Sciences Volume 2*(No.1), Pages 1-7.

Fisher, W.P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21(1)*, 1095.

John, M. L. (1994). Sample size and item calibration (or person measure) stability. *Rasch Measurement Transactions*, 7(4), 328. https://www.rasch.org/rmt/rmt74m.htm.

Kakia, L., Popov, H. N., & Arani, A. M. (2016). A study of relationships between parents' and teachers' demographic factors and their judgments about children's activities and school readiness in primary schools of Tehran. *International Journal of Comparative Education and Development*, *18*(2), 100-119. https://doi.org/10.1108/IJCED-10-2015-0006

Lim, S.M., Rodger, S., & Brown, T. (2009). Using Rasch analysis to establish the construct validity of rehabilitation assessment tools. *International Journal of Therapy and Rehabilitation, 16(5), 251–260.*

Linacre, J.M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3(1), 85–106.*

Linacre, J.M. (2019a). *A user's guide to Winsteps Ministeps Rasch-model computer programs*. Beaverton, Oregon: Winsteps.com.

Linacre, J.M. (2019b). *Winsteps® (Version 4.4.1)* [Computer software]. Beaverton, OR. Retrieved from winsteps.com.

Lynch, D., Smith, R., Yeigh, T., & Provost, S. (2019). A study into "organisational readiness" and its impacts on school improvement. *International Journal of Educational Management*.

Ministry of Education Malaysia. (2018a). *Pemansuhan amalan peperiksaan pertengahan dan akhir tahun murid tahap 1 (tahun 1, tahun 2 dan tahun 3) sekolah rendah Kementerian Pendidikan Malaysia mulai tahun 2019*. WP Putrajaya.

Moosa, V., Shareefa, M., Adams, D., & Mohamed, A. (2020a). Assessing Teacher Readiness for Inclusive Education in Maldives: Instrument Modification and Validation. *International Journal of Disability, Development and Education*, 1-17. https://doi.org/10.1080/1034912x.2020.1837355.

Myint Swe, K. (2020). *Rasch Measurement : Applications in Quantitative Educational Research* [Book]. Springer.https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsebk&AN=2345201&site=eds-live.

Rorlinda Yusof, & Faridah Anum Abdul Wahid. (2014). Pengetahuan pentaksiran berasaskan sekolah dalam kalangan guru pelatih PISMP di Institusi Pendidikan Guru Kampus Ilmu Khas. *Jurnal Penyelidikan TEMPAWAN*, *Volume 31*, Pages 1-10.

Rohaya Talib, & Mohd Najib Abd Ghafar. (2008). Pembinaan dan pengesahan instrumen bagi mengukur tahap literasi pentaksiran guru sekolah menengah di malaysia. *Proceeding of Seminar Penyelidikan.*

Smith, B.W., Epstein, E.M., Ortiz, J.A., Christopher, P.J., & Tooley, E.M. (2013). The foundations of resilience: What are the critical resources for bouncing back from stress? In S. Prince-Embury & D.H. Saklofske (Eds.), *Resilience in children, adolescents, and adults (pp. 167–187).* New York, NY: Springer. doi:10.1007/978-1-4614-4939-3_13.

Smith, B.W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15(3),* 194–200.

Suah, S. L., Ong, S. L., & Shuki Osman. (2010). Pentaksiran pembelajaran pelajar: Amalan guru-guru di Malaysia. *Akademia*.

Wright, B.D., & Masters, G.N. (1982*). Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B.D., & Stone, M.H. (1999). *Measurement essentials*. Wilmington, DE: Wide Range Inc.

Wright, B.D., & Masters, G.N. (1982*). Rating scale analysis*. Chicago, IL: MESA Press.

Wolfe, E.W., & Smith, E.V. (2007). Instrument development tools and activities for measure validation using Rasch models: Validation activities. *Journal of Applied Measurement, 8(2), 204–234.*

Wong, N., Rindfleisch, A., & Burroughs, J.E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research, 30(1), 72–91.*

Zahari Suppian. (2014). *Pembinaan dan pengesahan instrumen untuk mengukur tahap kompetensi guru pelatih dalam pentaksiran bilik darjah* [Doctoral Thesis, Universiti Kebangsaan Malaysia]. Bangi, Malaysia.https://books.google.com.my/books?id=bhSLAQAACAAJ.