Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2023, Volume 8, Issue 12, e002618
DOI: https://doi.org/10.47405/mjssh.v8i12.2618

RESEARCH ARTICLE

Field Expert Assessment on the Inventory of Science Teacher Readiness in Implementing Classroom-based Assessment

Mat Rasid Ishak[1*], Hidayah Mohd Fadzil[2], Harris Shah Abd Hamid[3]

[1]Faculty of Education, University of Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia.
Email: matrasid.ishak@gmail.com
[2]Faculty of Education, University of Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia.
Email: hidayahfadzil@um.edu.my
[3]Faculty of Management, Education and Humanities, University College MAIWP International, Kuala Lumpur, Malaysia.
Email: drharris@ucmi.edu.my

CORRESPONDING AUTHOR (*):
Mat Rasid Ishak
(matrasid.ishak@gmail.com)

KEYWORDS:
Classroom-Based Assessment
Science Teacher Readiness
Multi Facet Rasch Model
Inventory validity

CITATION:
Mat Rasid Ishak, Hidayah Mohd Fadzil & Harris Shah Abd Hamid. (2023). Field Expert Assessment on the Inventory of Science Teacher Readiness in Implementing Classroom-based Assessment. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 8*(12), e002618. https://doi.org/10.47405/mjssh.v8i12.2618

ABSTRACT

Teacher readiness is the willingness of a teacher to implement a programmed Classroom-Based Assessment (CBA) successfully, which is crucial to ensure that a program can be implemented at the individual or organizational level. The readiness of science teachers to implement CBA determines the direction and success of primary school assessments in Malaysia. However, there are limitations in measuring science teachers' readiness to implement CBA. Therefore, there is a need to develop an instrument that can measure the ability. This study was conducted to validate the Science Teacher Readiness Inventory in CBA (ISTRI-CBA) using a Multi-Faceted Rasch Measurement Model (MFRM). A sample of five officers from the Ministry of Education, Malaysia, served as a panel of field experts in assessments, science education, psychometrics, psychology, and language. Based on applying the MFRM, 162 items were retained from an initial pool of 185 items. The panel of field experts was good, with all five having a level of leniency and severity not too far when rating the item across the five dimensions. Expert evaluation using four-point rating scale criteria was used for clarity, accuracy, reasonableness, and brevity. The Results show ISTRI-CBA for the constructs of CBA knowledge, CBA skills, Attitudes, and Professionalism Values were considered unidimensional, while the resource support construct was multidimensional. Overall, the item reliability was satisfactory, ranging between 0.70 to 0.87 and the chi-square value (0.00) was significant, thus demonstrating that the items contained in the Science Teacher Readiness Inventory in the Classroom-Based Assessment dimension have good internal consistency.

Contribution/Originality: The paper's primary contribution is finding that unique psychometric properties with MFRM analysis for Inventory Science Teacher Readiness implementing Classroom-Based Assessment have good internal consistency. This study

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

has also proven that four of the five constructs ISTRI-CBA, namely CBA knowledge, CBA skills, attitudes, and values of professionalism, are unidimensionality.

1. Introduction

This study focuses on the readiness of science teachers to implement Classroom Based Assessment (CBD) at primary schools in Malaysia. It begins with an explanation of the science teacher readiness in CBA and how to develop constructs or dimensions in the readiness of science teachers. To unravel the constructs representing the readiness of science teachers, this study examined and analyzed domestic and foreign studies to reveal the proposed constructs and terms of readiness. The primary research and analysis of the previous study aim to identify the primary constructs that measure teachers' readiness in assessment and match it with the cultural context and curriculum in Malaysia.

Readiness is when a person is fully prepared mentally and physically to successfully implement a program (Dewan Bahasa & Pustaka, 2017). Readiness also refers to a condition where an individual has mastered the skills, knowledge, and opportunity to understand any new cognitive activity (Meisels, 1998). However, the use of constructs in measuring teacher readiness, especially for science subjects, is different for each country as the content and curriculum assessed differ according to the education model in that country.

According to the theory of individual readiness by Madsen et al. (2006), there are three dimensions of readiness: experiential, environmental, and affective. Therefore, the teacher's readiness in this study is specific to assessment, compared to five existing assessment models in previous studies that have been analyzed to determine the dimensions that coincide in measuring science teachers' readiness to implement CBA. The five existing assessment models used and analyzed include the Teacher Concept of Assessment (Brown, 2013), 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 & Mohd, 2008) and Model of Teacher Assessment Practice Inventory (IAPG) (Suah & Ong, 2012).

The results of the analysis discovered five appropriate dimensions to measure the readiness of science teachers in CBA: science teacher knowledge in CBA, skills in implementing CBA, resource support (infrastructure and infostructures such as training and teaching aids), attitudes, and values of professionalism. The result of the adaptation of Individual Readiness Theory with five assessment models was later translated into the theoretical framework of this study, as shown in Figure 1.

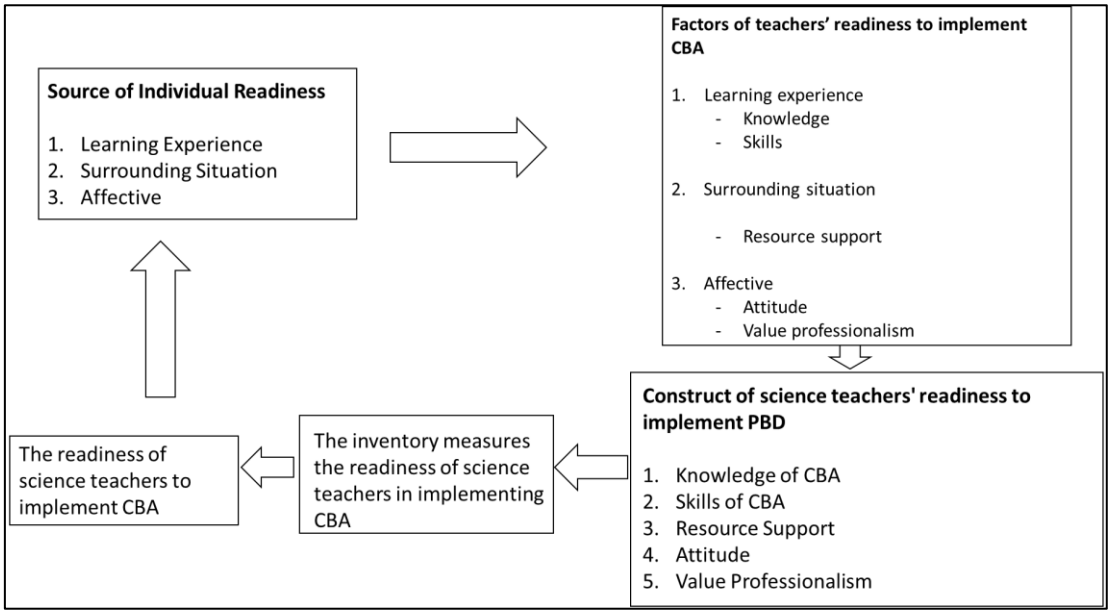
1.1. Research Objectives

For this study, two research objectives were determined to achieve the desired goal, namely;

i. Develop and verify ISTRI-CBA Inventory items based on the assessment of the CBA Field Expert Panel.
ii. Test the quality characteristics of ISTRI-CBA items using the Multi-Facet Rasch Model (MFRM).

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Figure 1: Theoretical Framework for ISTRI-CBA



2. Literature Review

2.1 Classroom-Based Assessment

Brown (2013) states that among the main concepts that teachers should master in assessment is the appropriateness of the assessment used. The suitability of this assessment method can be obtained if the teacher is fully prepared and understands the assessment concept. *Assessment* is a learning process that involves recording, storing, measuring, and summarizing student achievement levels (Opre, 2015). In assessment, there are two assessment methods, namely formative and summative. These two methods can be used for student assessment in the classroom. Formative assessment is continuously assessing and monitoring student performance throughout the year. At the same time, a summative assessment is a formal assessment done at the end of a learning topic to assess student achievement (Azizi & Kamisah, 2018). Assessment in science subjects involves formative and summative assessments during and after Teaching and Learning (P&P). In Malaysia, the assessment system used by the Ministry of Education Malaysia (MoE) in science subjects today is School Based Assessment (SBA) and Classroom Based Assessment (CBA) (Examination Board, 2014). SBA is for students from Standard 4 until Standard 6 and Form 1 to Form 3. In contrast, CBA is for students from Standard 1 until Standard 3, as well as Grade 1 to Grade 3.

Assessment is a formal procedure that involves collecting and using information to determine student status to allow teachers to decide on student abilities concerning their Teaching and Learning (T&L) status. The CBA implemented in 2019 involved only level-one primary school students and is a process to obtain student development, progress, abilities, and mastery data to achieve classroom learning objectives (Ministry of Education Malaysia, 2018). Curriculum Development Department (2019) asserted that the implementation of CBA is used to obtain information about student progress and analyze it to make improvements. The process of getting this information is a form of continuous feedback towards improving assessment and student development. The implementation of CBA is done according to a method appropriate for the teachers in the class so that it can affect the mastery of science learning on the level of student

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

achievement. CBA is an assessment fully implemented by science subject teachers, including teaching planning, assessment, marking, recording, and reporting (Zahari & Jamil, 2014).

The content in the science subject's curriculum aims to foster students ' creative and exciting learning. The content of the science curriculum is also for students to master science concepts, thinking skills, scientific skills, scientific attitudes, and moral values based on learning based on experience, discovery, and investigation (Ministry of Education Malaysia, 2017b). Therefore, CBA implementation in science subjects follows the curriculum standard used in the science T&L known as the 2017 Revised Primary School Standard Curriculum, which was specially developed for only science subjects so that students can fully master the set achievement standards. The science curriculum content standards are designed to guide the development of students based on the performance standards set. Student performance standards in the Curriculum and Assessment Standard Document are divided into four main assessment components: scientific knowledge, scientific skills, scientific attitudes, and moral values. These four components are assessed and scored using the Proficiency Level 1 (PL1) to Proficiency Level 6 (PL6) rating scale. Assessment, in general, is also done throughout the student's T&L in the class, including the assessment of science process skills, which is done through the teacher's professional judgment (Ministry of Education Malaysia, 2017a). PL assessment criteria are explained in Table 1 for Performance Level of Scientific Knowledge and Skills, Table 2 for Performance Level Translation for Scientific Attitudes and Pure Values, and Table 3 for Elementary School Science General Performance Level Statement, demonstrating the interpretation of CBA assessment in science subjects.

Table 1: Performance Level of Scientific Knowledge and Skills

Performance Level	Interpretation
1	Pupils can recall learning through scientific knowledge and basic scientific skills applied.
2	Pupils can re-understand learning through scientific knowledge and basic scientific skills and explain the understanding of the T&L to the teacher.
3	Pupils can re-apply learning through scientific knowledge and basic scientific skills to perform simple tasks given by the teacher in class.
4	Pupils can reanalyse learning through scientific knowledge and basic scientific skills in problem-solving.
5	Pupils can re-evaluate learning through scientific knowledge and basic scientific skills through problem-solving aspects and make decisions in tasks requested by the teacher.
6	Pupils can recreate learning through scientific knowledge and basic scientific skills in problem-solving aspects. They can make decisions or be able to perform tasks in various new situations, or be able to do tasks creatively and innovatively.

Table 2: Performance Level Translation for Scientific Attitudes and Pure Values

Performance Level	Interpretation
1	Students have interest in science subjects.
2	Students have an interest in and curiosity about science.
3	Students are interested, curious, honest, and accurate about science.

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

4	We are recording science data.
5	Students have interest, and curiosity, be honest and accurate in recording science data and dare to try and be systematic.
6	Pupils have interest, and curiosity, be honest and accurate in recording science data, dare to try, be systematic, cooperate, diligent, and persevere in carrying out the tasks given by the teacher in the class.

Table 3: Elementary School Science General Performance Level Statement

Performance Level	Interpretation
1	Pupils know about essential matters related to science and science skills and show interest in T&L science or can implement basic skills of science concepts or give good responses in essential matters related to science for a topic taught in class.
2	Pupils can respond by understanding related science knowledge and skills and demonstrating.
3	An understanding of the science taught where students could explain a topic taught in the form of communication and show a deep sense of curiosity.
4	Pupils can apply related science knowledge, and skills learned in class to carry out simple tasks honestly and accurately record science data for tasks in a situation in the classroom.
5	Pupils can analyse related science knowledge, and science skills learned in T&L in the context of systematic problem solving and dare to try something new.
6	Pupils can assess related science knowledge and skills to implement a science process skill on assignments for problem-solving situations and make decisions according to procedures in a systematic, analytical, cooperative, diligent, and persevering manner while in class.

Implementing CBA depends on science teachers' ability to understand and relate the knowledge and skills they have for assessment (Khalil & Awang, 2016). Mastering assessment through direct training can improve teacher skills, which can be later translated with T&L in the classroom to make practical assessments of student achievement (Koloi-Keaikitse, 2016). Accordingly, teachers must master, understand, practice, and appreciate the ways of implementing CBA in science subjects to ensure the success of CBA implementation and further help to effectively measure student development so that it can be translated in line with the requirements of the National Education Philosophy, which demands a holistic student development in the classroom.

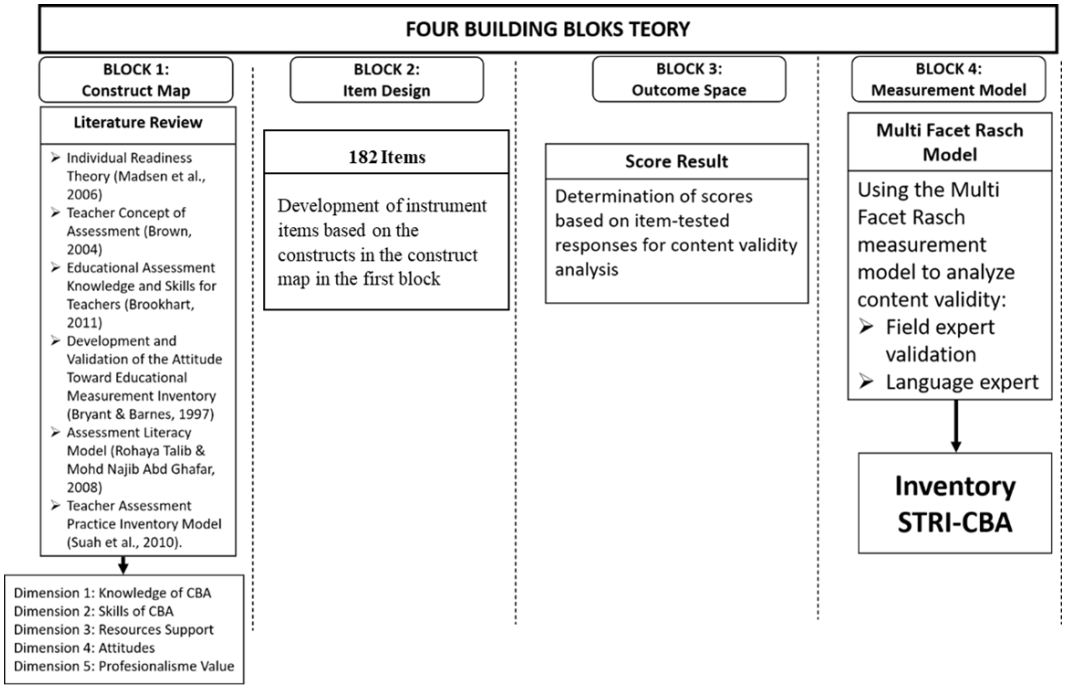
2.2. ISTRI-CBA Inventory Development

The development of the ISTRI-CBA inventory in this study utilized the conceptual framework adapted from the Four Building Blocks Theory by Wilson (2005), Individual Readiness Theory by Madsen et al. (2006), and the analysis of five assessment models. Figure 2 shows inventory development from the first block to form science teacher readiness constructs. The second block involves the development of 182 ISTRI-CBA items based on the construct map in the first block. Next, in the third block, the results of item test scores for content validity are analyzed using MFRM. Finally, in the fourth and last block, items are tested using the Multi Facet Rasch Model to confirm that the item's validity can be accepted, refined, or dropped based on the psychometric characteristics set in the Rasch Model. The analysis of selecting ISTRI-CBA items with psychometrics that

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

coincide in measuring the willingness of science teachers to implement CBA can prove the validity of the measurement model produced (Rasch, 1993).

Figure 2: Instrument Development Model Four Building Blocks Theory by Wilson (2005)



3. Research Methods

This research is a quantitative approach because the researcher used a questionnaire survey in this study using quantitative data. This study uses a MFRM method to collect data on the field expert panel related to classroom-based assessment practices. Five officers were purposively chosen from the Ministry of Education, Malaysia and given self-administered questionnaires.

3.1. Participant

Five experts in the field of science assessment at the Malaysian Ministry of Education participated in the evaluation of this inventory. The study sample consisted of various areas and backgrounds selected by purposive sampling based on their expertise. Table 4 below illustrates the characteristics of the panel of field experts involved in the ISTRI-CBA evaluation.

Table 4: List of Field Expert Panels in ISTRI-CBA Validation

Expert	Title	Gander	Position	Department	Expertise
Field expert 1	Dr.	Men	JUSA B	IPG	CBA
Field expert 2	Mr.	Men	Assistant Director	*Nazir*	CBA
Field expert 3	Mrs.	Women	Head Master	School	Pengurusan
Field expert 4	Mrs.	Women	Head Science Panel	Science Teacher	SME in Science
Field expert 5	Mrs.	Women	Counselling Teacher	School	Psychology

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

3.2. Measure

This study employed the ISTRI-CBA Instrument, which consisted of a four-point Likert rating scale from 1 (Strongly Disagree) to 4 (Strongly Agree). The content validity of ISTRI-CBA suggests internal consistency reliability (Cronbach's alpha) ranging from 0.70 to 0.91 (Khairul Azhar et al., 2018).

3.3. Procedure

The researcher has officially applied to the five expert panels for their approval as ISTRI-CBA assessors. Agreeable responses to each expert panel are critical to the researcher. After getting the consent of all the experts selected based on their expertise, letters of appointment as the ISTRI-CBA assessment panel from the University of Malaya were issued to the panels involved. After completing documentation and approval, the researcher first sent the Inventory to be evaluated before a meeting was held with each expert panel to assess ISTRI-CBA. The meeting with the expert panels took place online and face-to-face, depending on the suitability of the time and the COVID-19 factor at that time.

3.4. Data analysis

Rasch analysis was performed on the data of this study using Minifac (Version No. 3.83.6) (Thomas, 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). According to the recommendations of Lim, Rodger, and Brown (2009), an excellent study to test the psychometric properties of the inventory by looking at the main assumptions in the Rasch model for content validity, response process, and internal structure as outlined in the book Rasch Measurement: Application in Quantitative Educational Research namely unidimensionality, measurement scale functionality, item properties, person response, item targeting, and reliability (Myint Swe, 2020).

3.5. Data Analysis

All study data were analyzed using statistical methods with MINIFAC software. This study employed Multi Facet Rasch Model (MFRM) analysis and an expert assessment review form with three facets (Item, Aitem and Rater).

4. Results

4.1. Field Expert Evaluation Findings

Responses from professional evaluators were analyzed using a Multi Facet Rasch Model (MFRM) through Minifac software, an application developed to analyze measurement instruments involving a panel of assessors with more than one person. The MFRM was preferred for its ability to identify unexpected responses from experts and inappropriate judgments and predict missing data based on systematic response patterns (Thomas, 2015). A multi-faceted Rasch analysis conducted on the ratings by a panel of professional experts provides information on the reliability and separation index of the raters, the reliability and separation index of the items, the inventory rating criteria on each item, and the unexpected responses of the raters. Findings in MFRM for logit measurement

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

were arranged according to the evaluator's difficulty level to agree with the item evaluated based on the criteria set. The higher (positive) the measurement logit value, the more difficult it is for the evaluator panel to agree with the item being evaluated, while the lower the measurement logit value, the easier it is for the evaluator panel to agree with the item being assessed (Wen-Ta et al., 2018). The MFRM analysis in Table 5 shows the findings on the assessor's severity in assessing the ISTRI-CBA items.

Table 5: Judgment Summary of Field Expert for STRI-CBA Inventory

Matters	Knowledge of CBA	Skills of CBA	Resource Support	Attitude	Value of Professionalism
Reliability	0.98	0.98	0.90	0.98	0.88
Separation	7.16	6.97	2.99	6.94	2.66
Outfit MnSq	0.52 until 5.97	0.75 until 1.20	0.68 until 1.84 (D- Min)	0.95 until 1.61	0.94 until 1.09 (D- Min)
Measure (assessor's assertiveness)	-0.98 until -6.28	-0.76 until -3.86	-1.46 until -8.32	-0.39 until -3.89	-0.71 until -6.99
Rater exact agreement	44.6%	44.9%	46.2%	42.2%	32.7%

Table 5 shows the findings of the separation value for the rigor of the field expert's assessment for the construct of CBA knowledge, CBA skills, resource support, attitude, and professionalism value between 2.66 to 7.16 with the rater reliability index being in the excellent category between 0.88 to 0.98 with a chi-square value (K) Significant p=0.00. Accordingly, the findings proved a significant difference in the strictness of the assessors in giving assessments on the items found in the ISTRI-CBA.

Based on the logit value in Table 5 for the CBA knowledge construct, it was found that Expert B was the most assertive and Expert D was the easiest to agree to the items contained in the ISTRI-CBA. The consistency of the evaluator's strictness can be determined through the value of Infit MnSq or Outfit MnSq received between the logit value 0.5 to 1.5 (George & Stefanie, 2018). It was observed that the Outfit values for experts B, A, E, and C were consistent with the model (item fit), which is consistent in giving scoring evaluation to the items. At the same time, expert D has an Outfit value that did not match the model in scoring. Based on the findings of expert rigor in Table 5, there were three groups of experts with different levels of assessment difficulty. The group consisted of Expert groups A and B, who agreed on the items with a logit between -0.98 and -1.49. The second Expert Group consisted of C and E, with difficulty logit values between -2.43 and -3.38. At the same time, the third group, Expert D, has a difficulty logit value of -6.28. These results indicate that there were groups of experts with the same and different views related to the items in the CBA knowledge construct even though no meeting or discussion took place between all the expert panels involved in the assessment of this Inventory. Overall, the rater's agreement percentage in this construct was 44.6% compared to the model's expected value of 49.1%. The very close percentage difference between the actual value and the model shows the Expert's ability to evaluate ISTRI-CBA excellently.

The findings of the assessor's firmness for the CBA Skill construct in Table 5 showed that Experts A and B were the most firm and Expert D was the easiest to agree to the ISTRI-CBA items. The consistency of the evaluator's strictness determined through the value of

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Infit MnSq or Outfit MnSq found that all experts in this construct were consistent with a consistent model in giving scoring evaluations to the items. The findings for this construct also showed three groups of experts with different levels of assessment difficulty for the CBA skill construct. The group consisted of Expert groups A and B, who agreed on the items in the ISTRI-CBA knowledge construct with a logit of -0.76. The second Expert Group consisted of E and C, with difficulty logit values from -2.28 to -2.76. At the same time, the third group comprised expert D, who has a difficulty logit of -3.86. This result showed that three groups of expert panels have the same views related to the items contained in the CBA skills construct even though there was no meeting between all the expert panels involved in evaluating this Inventory. Overall, the rater's agreement percentage in this construct was 44.9% compared to the Model's expected value of 45.9%. This very close percentage demonstrated that the expert panel's ability to evaluate ISTRI-CBA was excellent when the actual value was relative to the expected Model value.

Next, the content validity for the source support construct involved the assertiveness of the assessor in this study. Based on the logit value in Table 5, it was found that Experts B and E were the strictest, while Expert D was the easiest to agree to the items in this construct. The consistency of the evaluator's strictness found through Outfit MnSq for four experts, namely A, B, C, and E, was consistent with the model based on the evaluation findings performed on the ISTRI-CBA items. There were three groups of experts with different levels of assessment difficulty. The group consisted of Expert groups B and E, who agreed on the items in the resource support construct with a logit of -1.46 to -1.54. The second Expert Group consists of A and C, with difficulty logit values from -2.07 to -2.47. The third group, D, has a difficulty logit value of -2.07. This result suggested that the three groups of expert panels have the same views related to the items contained in the resource support construct, even though there was no meeting or discussion between the expert panels involved in evaluating this Inventory. Overall, the rater's agreement percentage in this construct was 46.2% compared to the model's expected value of 51.5%. The difference in the percentage of evaluator agreement showed the ability of the actual value to approach the model value with good conditions.

The content validity for the attitude construct also found the assertiveness of the assessors in this study based on Table 5, showing Expert A as the most assertive and Expert D as the easiest to agree to the items contained in the ISTRI-CBA construct. The consistency of accuracy of the Outfit MnSq raters for four of the five experts was consistent with the model being compatible in assigning scoring ratings to the items. Furthermore, the findings revealed four groups of experts with different levels of assessment difficulty. The group consisted of Expert A, who agreed on the items in the ISTRI-CBA attitude construct with a logit of -0.39. The second Expert Group consisted of B and E, with difficulty logit values from -1.09 to -1.26. The third group, which was C, has a difficulty logit in the range of -1.73. While the fourth group, D, has a difficulty logit of -3.89. This result shows that four expert panels in the evaluation of this Inventory have the same view related to the items contained in the attitude construct. Findings related to the rater's agreement percentage in this construct was 42.2% compared to the model's expected value, which was 47.5%. This percentage difference shows the evaluator's ability to evaluate ISTRI-CBA very well compared to the expected percentage value of the model.

Finally, the content validity analysis for the value construct of professionalism in Table 5 shows the strictness of the evaluator. Expert A was the strictest, whereas Expert E was the easiest to agree to the items contained in the ISTRI-CBA. The consistency of the

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

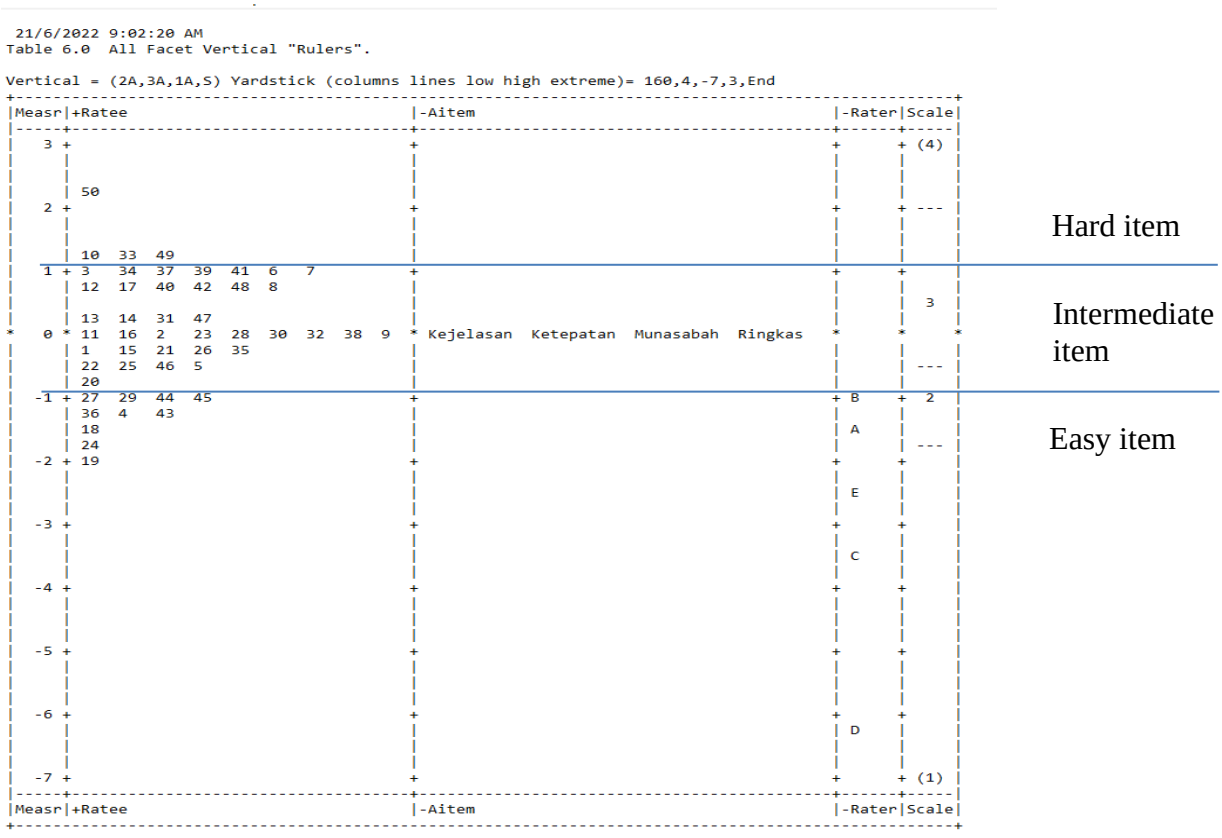
evaluator's strictness showed that four out of five experts were consistent with the consistent model in giving scoring evaluations to the items. Table 5 also shows that there were two groups of experts with different levels of assessment difficulty. The group consisted of Expert groups A, B, C, and E, who agreed on the items in the attitude construct with a logit between -0.71 and -1.29. At the same time, the second expert group was D, which has a logit difficulty value of -6.99. This result shows that two groups of expert panels in the evaluation of this Inventory have the same view related to the items contained in the professionalism value construct. The percentage of evaluator agreement in this construct was 32.7% compared to the model's expected value of 40.1%. This close percentage of difference shows the ability of the actual value to approximate the model value with perfect conditions for the evaluation made by the expert panels.

4.2. Validation Item ISTRI-CBA

4.2.1 CBA Knowledge Construct

The findings of the MFRM analysis of the CBA knowledge construct map in Figure 3 displayed the item difficulty, item evaluation level, and rater evaluation level (expert panel). The interpretation of this construct map showed that the CBA knowledge construct items have content validity that has met the MFRM model and was acceptable. This result was demonstrated by breaking item categories into complex, medium, and accessible. However, items below the logit value of -2 show that the item is too easy and less suitable to measure the CBA knowledge construct. Therefore, item 19 was proposed to be improved or dropped. The quality of the items shown in Figure 1 shows that the higher the item on the logit value, the better the items for ISTRI-CBA in the MFRM Measurement Model (George & Stefanie, 2018).

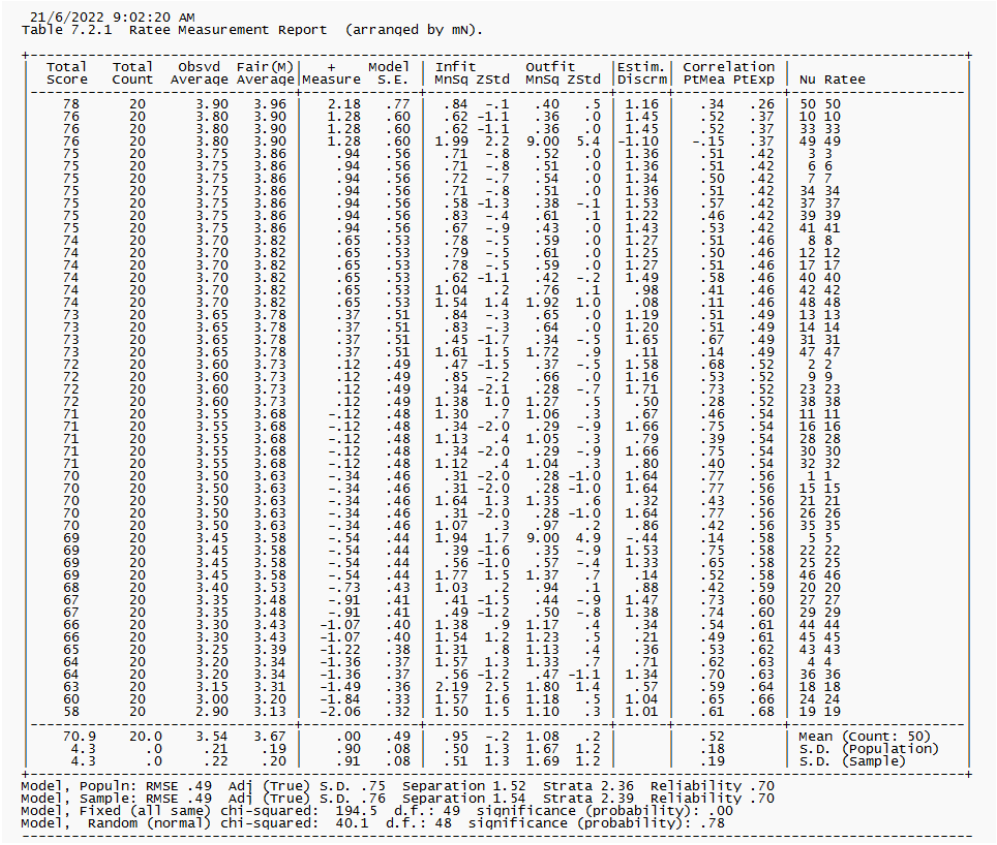
Figure 3: CBA Knowledge Construct Map



© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Next, the items were examined for content validity in the MFRM analysis by looking at the statistical compatibility of the items based on the Outfit MnSq value (0.5 to 1.5) or Outfit Zstd (-2 to +2) or the Point Measure Correlation value (0.4 to 0.85) (George & Stefanie, 2018). The findings of Outfit MnSq and Outfit Zstd shown in Table 3 summarize the overall quality of the CBA Knowledge construct items based on Figure 4. It was found that items 5 and 49 should be dropped since they have values outside the set statistical compatibility. Appropriate items indicate that they can measure the same construct, while inappropriate items must be removed or improved (Villalonga-Olives et al., 2021). Findings showed the reading of Outfit MnSq for CBA knowledge construct items between 0.28 to 9.00, which will be dropped or refined for items, not between the value range of 0.5 to 1.5. The Outfit Zstd value is also referenced if the Outfit MnSq value is not within the appropriate logit value for consideration to be refined or dropped.

Figure 4: Analysis Fit Item in MFRM for the CBA Knowledge Construct



21/6/2022 9:02:20 AM
Table 7.2.1 Ratee Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu	Ratee
78	20	3.90	3.96	2.18	.77	.84	-.1	.40	.5	1.16	.34	.26	50	50
76	20	3.80	3.90	1.28	.60	.62	-1.1	.36	.0	1.45	.52	.37	10	10
76	20	3.80	3.90	1.28	.60	.62	-1.1	.36	.0	1.45	.52	.37	33	33
76	20	3.80	3.90	1.28	.60	1.99	2.2	9.00	5.4	-1.10	-.15	.37	49	49
75	20	3.75	3.86	.94	.56	.71	-.8	.52	.0	1.36	.51	.42	3	3
75	20	3.75	3.86	.94	.56	.71	-.8	.51	.0	1.36	.51	.42	6	6
75	20	3.75	3.86	.94	.56	.72	-.7	.54	.0	1.34	.50	.42	7	7
75	20	3.75	3.86	.94	.56	.71	-.8	.51	.0	1.36	.51	.42	34	34
75	20	3.75	3.86	.94	.56	.58	-1.3	.38	-.1	1.53	.57	.42	37	37
75	20	3.75	3.86	.94	.56	.83	-.4	.61	.1	1.22	.46	.42	39	39
75	20	3.75	3.86	.94	.56	.67	-.9	.43	.0	1.43	.53	.42	41	41
74	20	3.70	3.82	.65	.53	.78	-.5	.59	.0	1.27	.51	.46	8	8
74	20	3.70	3.82	.65	.53	.79	-.5	.61	.0	1.25	.50	.46	12	12
74	20	3.70	3.82	.65	.53	.78	-.5	.59	.0	1.27	.51	.46	17	17
74	20	3.70	3.82	.65	.53	.62	-1.1	.42	-.2	1.49	.58	.46	40	40
74	20	3.70	3.82	.65	.53	1.04	.2	.76	.1	.98	.41	.46	42	42
74	20	3.70	3.82	.65	.53	1.54	1.4	1.92	1.0	.08	.11	.46	48	48
73	20	3.65	3.78	.37	.51	.84	-.3	.65	.0	1.19	.51	.49	13	13
73	20	3.65	3.78	.37	.51	.83	-.3	.64	.0	1.20	.51	.49	14	14
73	20	3.65	3.78	.37	.51	.45	-1.7	.34	-.5	1.65	.67	.49	31	31
73	20	3.65	3.78	.37	.51	1.61	1.5	1.72	.9	.11	.14	.49	47	47
72	20	3.60	3.73	.12	.49	.47	-1.5	.37	-.5	1.58	.68	.52	2	2
72	20	3.60	3.73	.12	.49	.85	-.2	.66	.0	1.16	.53	.52	9	9
72	20	3.60	3.73	.12	.49	.34	-2.1	.28	-.7	1.71	.73	.52	23	23
72	20	3.60	3.73	.12	.49	1.38	1.0	1.27	.5	.50	.28	.52	38	38
71	20	3.55	3.68	-.12	.48	1.30	.7	1.06	.3	.67	.46	.54	11	11
71	20	3.55	3.68	-.12	.48	.34	-2.0	.29	-.9	1.66	.75	.54	16	16
71	20	3.55	3.68	-.12	.48	1.13	.4	1.05	.3	.79	.39	.54	28	28
71	20	3.55	3.68	-.12	.48	.34	-2.0	.29	-.9	1.66	.75	.54	30	30
71	20	3.55	3.68	-.12	.48	1.12	.4	1.04	.3	.80	.40	.54	32	32
70	20	3.50	3.63	-.34	.46	.31	-2.0	.28	-1.0	1.64	.77	.56	1	1
70	20	3.50	3.63	-.34	.46	.31	-2.0	.28	-1.0	1.64	.77	.56	15	15
70	20	3.50	3.63	-.34	.46	1.64	1.3	1.35	.6	.32	.43	.56	21	21
70	20	3.50	3.63	-.34	.46	.31	-2.0	.28	-1.0	1.64	.77	.56	26	26
70	20	3.50	3.63	-.34	.46	1.07	.3	.97	.2	.86	.42	.56	35	35
69	20	3.45	3.58	-.54	.44	1.94	1.7	9.00	4.9	-.44	.14	.58	5	5
69	20	3.45	3.58	-.54	.44	.39	-1.6	.35	-.9	1.53	.75	.58	22	22
69	20	3.45	3.58	-.54	.44	.56	-1.0	.57	-.4	1.33	.65	.58	25	25
69	20	3.45	3.58	-.54	.44	1.77	1.5	1.37	.7	.14	.52	.58	46	46
68	20	3.40	3.53	-.73	.43	1.03	.2	.94	.1	.88	.42	.59	20	20
67	20	3.35	3.48	-.91	.41	.41	-1.5	.44	-.9	1.47	.73	.60	27	27
67	20	3.35	3.48	-.91	.41	.49	-1.2	.50	-.8	1.38	.74	.60	29	29
66	20	3.30	3.43	-1.07	.40	1.38	.9	1.17	.4	.34	.54	.61	44	44
66	20	3.30	3.43	-1.07	.40	1.54	1.2	1.23	.5	.21	.49	.61	45	45
65	20	3.25	3.39	-1.22	.38	1.31	.8	1.13	.4	.36	.53	.62	43	43
64	20	3.20	3.34	-1.36	.37	1.57	1.3	1.33	.7	.71	.62	.63	4	4
64	20	3.20	3.34	-1.36	.37	.56	-1.2	.47	-1.1	1.34	.70	.63	36	36
63	20	3.15	3.31	-1.49	.36	2.19	2.5	1.80	1.4	.57	.59	.64	18	18
60	20	3.00	3.20	-1.84	.33	1.57	1.6	1.18	.5	1.04	.65	.66	24	24
58	20	2.90	3.13	-2.06	.32	1.50	1.5	1.10	.3	1.01	.61	.68	19	19
70.9	20.0	3.54	3.67	.00	.49	.95	-.2	1.08	.2		.52		Mean (Count: 50)	
4.3	.0	.21	.19	.90	.08	.50	1.3	1.67	1.2		.18		S.D. (Population)	
4.3	.0	.22	.20	.91	.08	.51	1.3	1.69	1.2		.19		S.D. (Sample)	

Model, Populn: RMSE .49 Adj (True) S.D. .75 Separation 1.52 Strata 2.36 Reliability .70
Model, Sample: RMSE .49 Adj (True) S.D. .76 Separation 1.54 Strata 2.39 Reliability .70
Model, Fixed (all same) chi-squared: 194.5 d.f.: 49 significance (probability): .00
Model, Random (normal) chi-squared: 40.1 d.f.: 48 significance (probability): .78

Based on Table 6, seven items were categorized as very good, 21 were good, 20 were acceptable for improvement, and two items needed to be dropped. The result for the item separation value was 1.52, which was less than the value of 2, while the item reliability index was 0.70. The chi-square value was significant, k<0.001, which shows that the items contained in this CBA knowledge still have good internal consistency (Koskimäki et al., 2021).

Table 6: Summary of Statistical Fit Item Categories for CBA Knowledge

Item Category	Value of *Outfit* MnSq/*Outfit* Zstd	Item	Total Item

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

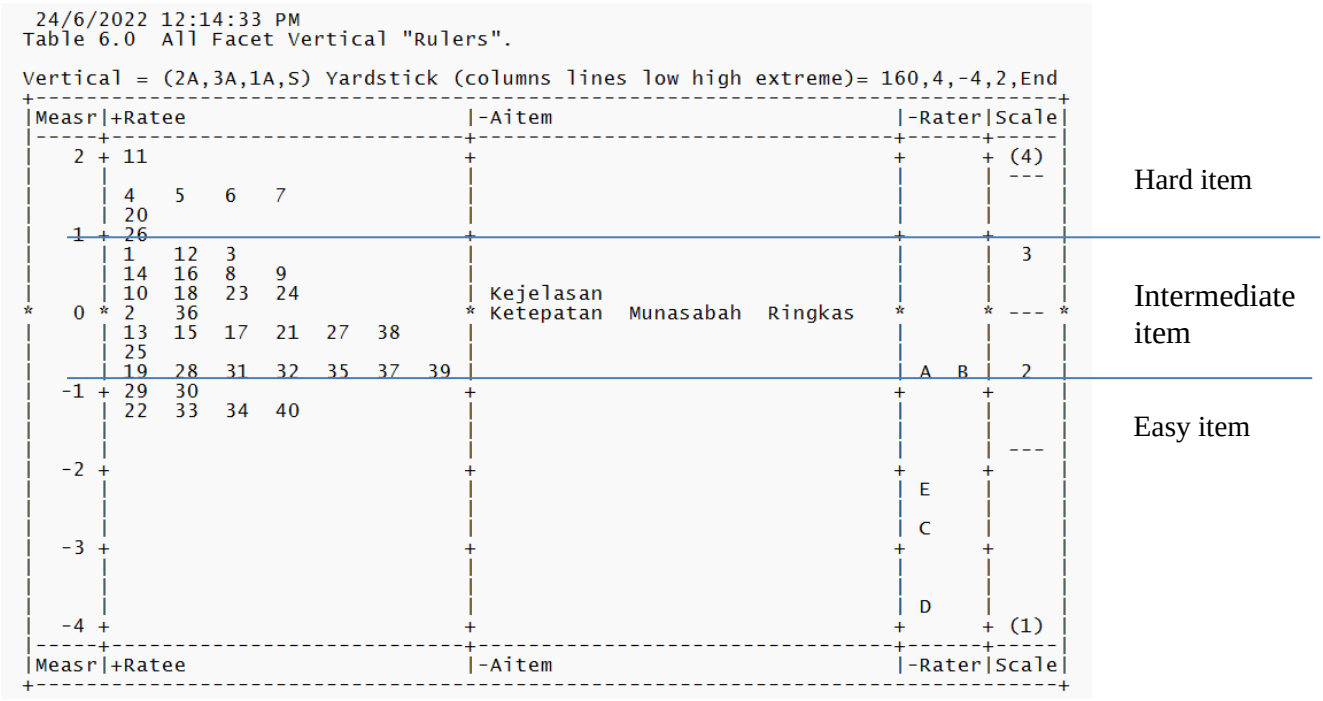
Very good	1.00	11, 19, 24, 28, 32, 43 and 44.	7
Good	Between 0.5 to 1.5	3, 4, 6, 7, 8, 9, 12, 13, 14, 17, 20, 21, 25, 29, 34, 35, 38, 39, 42, 45, 46.	21
Retained/ Modified	<0.5 and >1.5 but the Zstd value is still between -2 and +2	1, 2, 10, 15, 16, 18, 22, 23, 26, 27, 30, 31, 33, 36, 37, 40, 41, 47, 48 and 50.	20
Removed/dropped	<0.5 and >1.5 as well as value Zstd <-2 and >+2	5 and 49.	2

Mean = 0.00, S.D = 0.75, Separation = 1.52, Strata = 2.36, Reliability Index = 0.70
Chi-square = 194.5, d.f = 49, significant (probability) = 0.00

4.2.2. CBA Skill Construct

The following presents the findings of the MFRM analysis to test the content validity of the second construct, CBA skills. Figure 5 shows a construct map related to item difficulty, item evaluation level, and rater evaluation level. Findings showed the strengths of the items of complex, medium, and accessible categories. No item was found below the logit value of -2, indicating that no item was too easy in measuring the construct of CBA skills. The level of item-to-item evaluation was also found to measure all of them except for a few items that have a higher level of difficulty, namely items 4, 5, 6, 7, 20, and 11. Regarding the level of expert evaluation of items, experts can be divided into two simple and easy categories to give evaluation marks to items. One expert was too easy to provide an evaluation score, expert D, whereas no expert was seen as too difficult to provide a score. The aspect of item quality shown in Figure 5 shows that the higher the item's logit value, the better the items used for ISTRI-CBA in the MFRM Measurement Model (George & Stefanie, 2018).

Figure 5: Construct Map Skill of CBA



© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

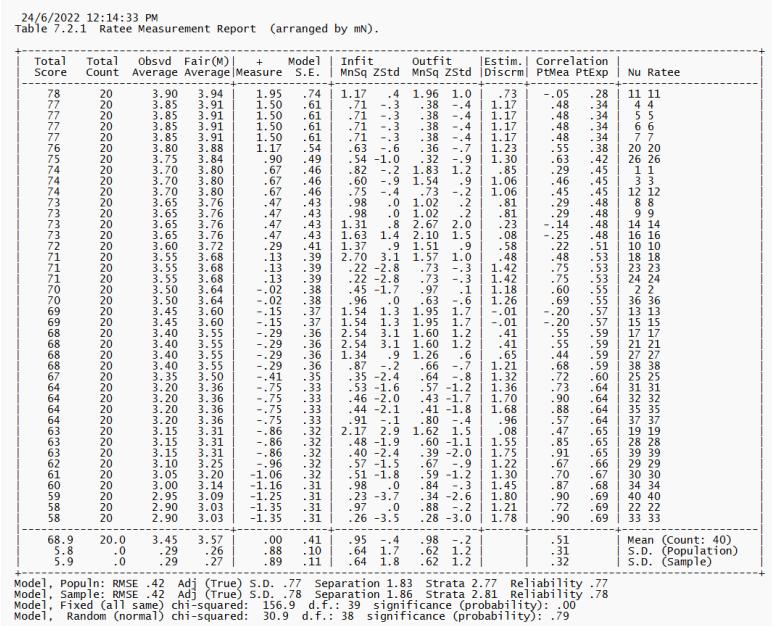
Further analysis also looked at statistical compatibility for instrument items based on Outfit MnSq (0.5 to 1.5) or Outfit Zstd (-2 to +2) or Point Measure Correlation (0.4 to 0.85) values (George & Stefanie, 2018). The results of Outfit MnSq and Outfit Zstd in Table 4 presented the overall summary of the quality of the CBA skill construct items. Some items do not fit and need to be dropped, namely items 14, 40, and 33, since they have inappropriate Outfit MnSq and Outfit Zstd values. Appropriate items indicate that they can measure the same construct, while inappropriate items must be removed or improved (Villalonga-Olives et al., 2021). The statistical compatibility of the CBA skill construct items can also be referred to in detail in the MFRM findings in Figure 6, which details whether to be discarded or retained. Based on Table 7, the item separation value was 1.83 less than the value of 2, while the item reliability index was 0.77. The chi-square value was significant, k<0.001, showing that the items in this CBA skill have good internal consistency. In conclusion, it was found that two items were categorized as very good, 17 items were categorized as good, 18 items could be accepted for improvement, and three items needed to be dropped.

Table 7: Summary of Statistical Fit Item Categories for CBA Skills

Item Category	**Value of *Outfit* MnSq/*Outfit* Zstd**	**Item**	**Total Item**
Very good	1.00	8 dan 9	2
Good	Between 0.5 to 1.5	2, 3, 10, 12, 22, 23, 24, 25, 27, 28, 29, 30, 31, 34, 36, 37 and 38	17
Retained/ Modified	<0.5 and >1.5 but the Zstd value is still between -2 and +2	1, 4, 5, 6, 7, 11, 13, 15, 16, 17, 18, 19, 20, 21, 26, 32, 35 and 39	18
Removed/dropped	<0.5 and >1.5 as well as value Zstd <-2 and >+2	14, 40 and 33	3

Mean= 0.00, S.D = 0.77
Separation = 1.83, Strata = 2.77, Reliability index = 0.77
Chi-square = 156.9, d.f = 39, significant (probability) = 0.00

Figure 6: Analysis Fit Item in MFRM for the CBA Skills Construct



24/6/2022 12:14:33 PM
Table 7.2.1 Ratee Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	PtExp	Nu Ratee
78	20	3.90	3.94	1.95	.74	1.17	.4	1.96	1.0	.73	-.05	.28	11 11
77	20	3.85	3.91	1.50	.61	.71	-.3	.38	-.4	1.17	.48	.34	4 4
77	20	3.85	3.91	1.50	.61	.71	-.3	.38	-.4	1.17	.48	.34	5 5
77	20	3.85	3.91	1.50	.61	.71	-.3	.38	-.4	1.17	.48	.34	6 6
77	20	3.85	3.91	1.50	.61	.71	-.3	.38	-.4	1.17	.48	.34	7 7
76	20	3.80	3.88	1.17	.54	.63	-.6	.36	-.7	1.23	.55	.38	20 20
75	20	3.75	3.84	.90	.49	.54	-1.0	.32	-.9	1.30	.63	.42	26 26
74	20	3.70	3.80	.67	.46	.82	-.2	1.83	1.2	.85	.29	.45	1 1
74	20	3.70	3.80	.67	.46	.60	-.9	1.54	.9	1.06	.46	.45	3 3
74	20	3.70	3.80	.67	.46	.75	-.4	.73	-.2	1.06	.45	.45	12 12
73	20	3.65	3.76	.47	.43	.98	.0	1.02	.2	.81	.29	.48	8 8
73	20	3.65	3.76	.47	.43	.98	.0	1.02	.2	.81	.29	.48	9 9
73	20	3.65	3.76	.47	.43	1.31	.8	2.67	2.0	.23	-.14	.48	14 14
73	20	3.65	3.76	.47	.43	1.63	1.4	2.10	1.5	.08	-.25	.48	16 16
72	20	3.60	3.72	.29	.41	1.37	.9	1.51	.9	.58	.22	.51	10 10
71	20	3.55	3.68	.13	.39	2.70	3.1	1.57	1.0	.48	.48	.53	18 18
71	20	3.55	3.68	.13	.39	.22	-2.8	.73	-.3	1.42	.75	.53	23 23
71	20	3.55	3.68	.13	.39	.22	-2.8	.73	-.3	1.42	.75	.53	24 24
70	20	3.50	3.64	-.02	.38	.45	-1.7	.97	.1	1.18	.60	.55	2 2
70	20	3.50	3.64	-.02	.38	.96	.0	.63	-.6	1.26	.69	.55	36 36
69	20	3.45	3.60	-.15	.37	1.54	1.3	1.95	1.7	-.01	-.20	.57	13 13
69	20	3.45	3.60	-.15	.37	1.54	1.3	1.95	1.7	-.01	-.20	.57	15 15
68	20	3.40	3.55	-.29	.36	2.54	3.1	1.60	1.2	.41	.55	.59	17 17
68	20	3.40	3.55	-.29	.36	2.54	3.1	1.60	1.2	.41	.55	.59	21 21
68	20	3.40	3.55	-.29	.36	1.34	.9	1.26	.6	.65	.44	.59	27 27
68	20	3.40	3.55	-.29	.36	.87	-.2	.66	-.7	1.21	.68	.59	38 38
67	20	3.35	3.50	-.41	.35	.35	-2.4	.64	-.8	1.32	.72	.60	25 25
64	20	3.20	3.36	-.75	.33	.53	-1.6	.57	-1.2	1.36	.73	.64	31 31
64	20	3.20	3.36	-.75	.33	.46	-2.0	.43	-1.7	1.70	.90	.64	32 32
64	20	3.20	3.36	-.75	.33	.44	-2.1	.41	-1.8	1.68	.88	.64	35 35
64	20	3.20	3.36	-.75	.33	.91	-.1	.80	-.4	.96	.57	.64	37 37
63	20	3.15	3.31	-.86	.32	2.17	2.9	1.62	1.5	.08	.47	.65	19 19
63	20	3.15	3.31	-.86	.32	.48	-1.9	.60	-1.1	1.55	.85	.65	28 28
63	20	3.15	3.31	-.86	.32	.40	-2.4	.39	-2.0	1.75	.91	.65	39 39
62	20	3.10	3.25	-.96	.32	.57	-1.5	.67	-.9	1.22	.67	.66	29 29
61	20	3.05	3.20	-1.06	.32	.51	-1.8	.59	-1.2	1.30	.70	.67	30 30
60	20	3.00	3.14	-1.16	.31	.98	.0	.84	-.3	1.45	.87	.68	34 34
59	20	2.95	3.09	-1.25	.31	.23	-3.7	.34	-2.6	1.80	.90	.69	40 40
58	20	2.90	3.03	-1.35	.31	.97	.0	.88	-.2	1.21	.72	.69	22 22
58	20	2.90	3.03	-1.35	.31	.26	-3.5	.28	-3.0	1.78	.90	.69	33 33
68.9	20.0	3.45	3.57	.00	.41	.95	-.4	.98	-.2		.51		Mean (Count: 40)
5.8	.0	.29	.26	.88	.10	.64	1.7	.62	1.2		.31		S.D. (Population)
5.9	.0	.29	.27	.89	.11	.64	1.8	.62	1.2		.32		S.D. (Sample)

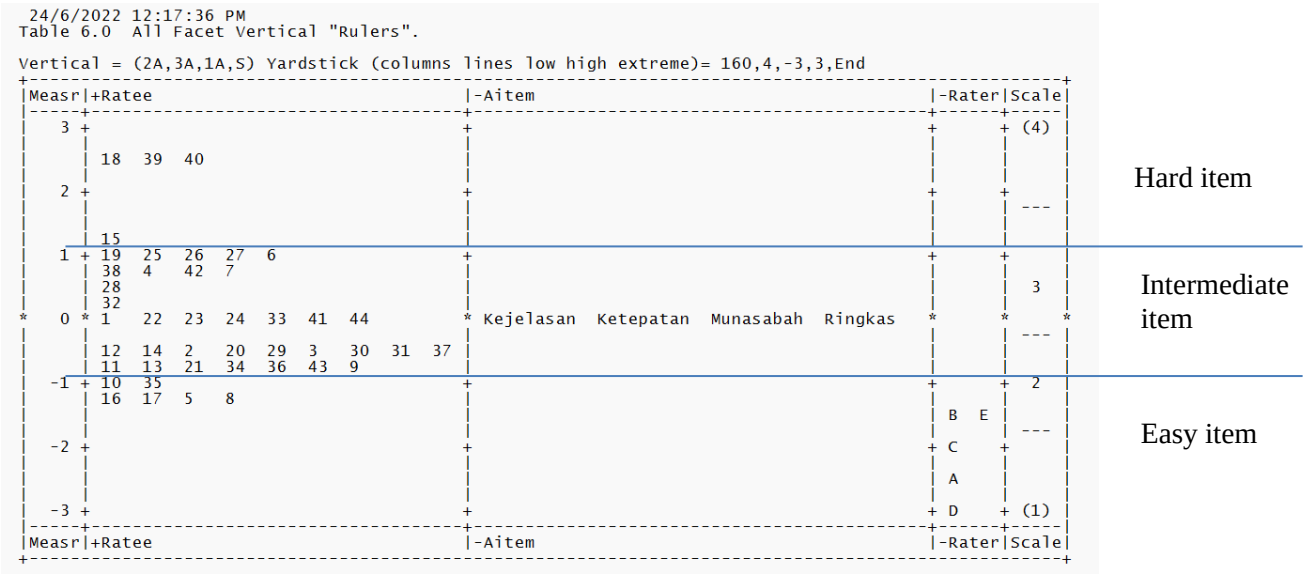
Model, Populn: RMSE .42 Adj (True) S.D. .77 Separation 1.83 Strata 2.77 Reliability .77
Model, Sample: RMSE .42 Adj (True) S.D. .78 Separation 1.86 Strata 2.81 Reliability .78
Model, Fixed (all same) chi-squared: 156.9 d.f.: 39 significance (probability): .00
Model, Random (normal) chi-squared: 30.9 d.f.: 38 significance (probability): .79

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

4.2.3 Resource Support Construct

MFRM analysis of content validity for the source support construct can be observed from Figure 7, which shows the construct map related to item difficulty, item evaluation level, and expert panel evaluation level divided into three categories of items: complex, medium, and accessible. Based on this finding, no item was seen below the logit value of -2, and the measurement of items against items was also found to be good and can measure the items as a whole. The results of expert evaluation of items can be divided into two categories: simple and easy to score. Experts B and E were simple raters in giving scores, while the other three experts were too easy in giving scores, namely experts A, C, and D. In terms of the quality of the items shown in the construct map in Figure 7, the higher the items in the value logit, the better the items used for ISTRI-CBA in the MFRM Measurement Model (George & Stefanie, 2018).

Figure 7: Construct Map for the Source Support Item



The analysis for the statistical compatibility of instrument items for resource support is shown in Table 8, which, based on Outfit MnSq and Zstd, found that some items need to be dropped, namely items 15, 2, 9, 11, 13, 21, 34, 36, 10, 16, 8 and 17. Appropriate items indicate that they can measure the same construct, while inappropriate items must be dropped or improved (Villalonga-Olives et al., 2021). The statistical compatibility of the items in the resource support construct can also be referred to in detail in Figure 8, which shows whether the items are to be discarded, retained, or refined.

Based on Table 8, the item separation value was 1.61 less than the value of 2, and the item reliability index was 0.72. The chi-square value was significant, k<0.001, which showed that the items contained in this resource support construct have a moderately good level of internal consistency and should be improved or dropped for items that are not appropriate. There were also expert comments on items with the same meaning in measuring the construct. One item was categorized as excellent, 20 items were categorized as good, 11 as acceptable for improvement, and 12 items to be dropped.

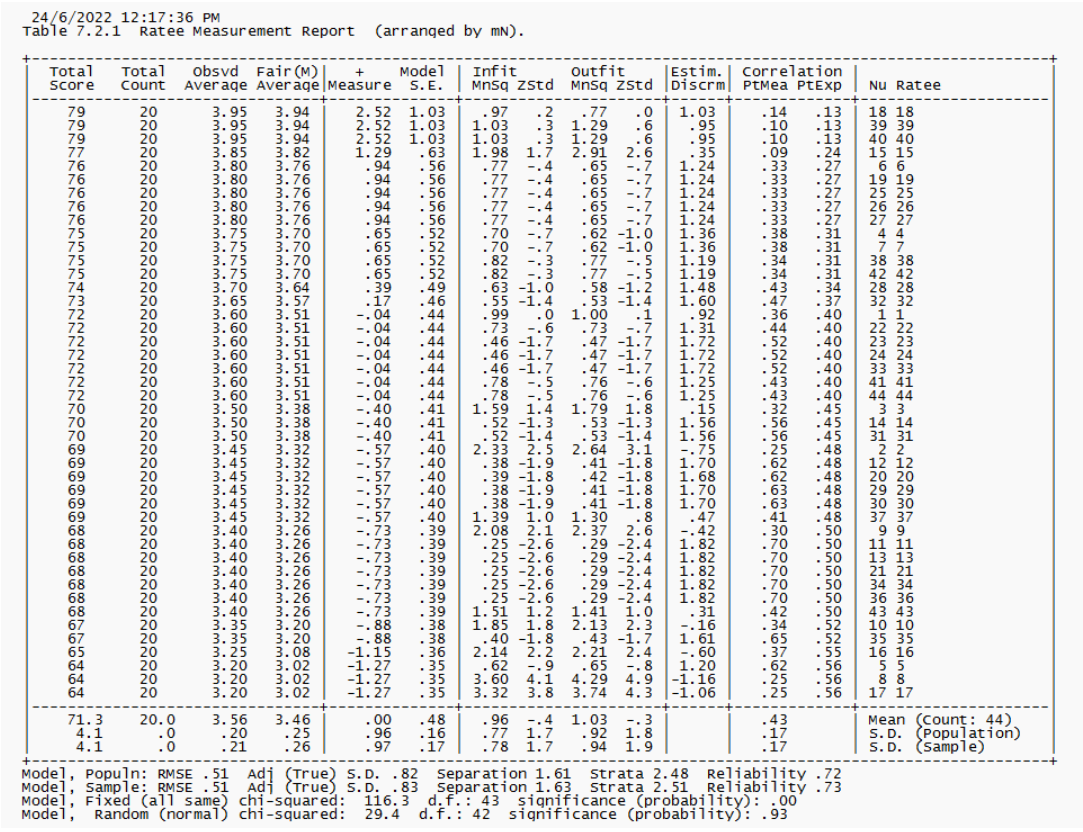
© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Table 8: Summary of Statistical Fit Item Categories for Resources Support Construct

Item Category	Value of *Outfit* MnSq/*Outfit* Zstd	Item	Total Item
Very good	1.00	1	1
Good	Between 0.5 to 1.5	18, 39, 40, 6, 19, 25, 26, 27, 4, 7, 38, 42, 28, 32, 22, 14, 31, 37, 43 and 5	20
Retained/ Modified	<0.5 and >1.5 but the Zstd value is still between -2 and +2	23, 24, 33, 41, 44, 3, 12, 20, 29, 30 and 35	11
Removed/dropped	<0.5 and >1.5 as well as value Zstd <-2 and >+2	15, 2, 9, 11, 13, 21, 34, 36, 10, 16, 8 and 17	12

Mean = 0.00, S. D = 0.82
Separation = 1.61, Strata = 2.48, Reliability index = 0.72
Chi-square = 116.3, d.f = 43, significant (probability) = 0.00

Figure 8: Analysis Fit Item in MFRM for the Resources Support Construct



24/6/2022 12:17:36 PM
Table 7.2.1 Ratee Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu	Ratee
79	20	3.95	3.94	2.52	1.03	.97	.2	.77	.0	1.03	.14	.13	18	18
79	20	3.95	3.94	2.52	1.03	1.03	.3	1.29	.6	.95	.10	.13	39	39
79	20	3.95	3.94	2.52	1.03	1.03	.3	1.29	.6	.95	.10	.13	40	40
77	20	3.85	3.82	1.29	.63	1.98	1.7	2.91	2.6	.35	.09	.24	15	15
76	20	3.80	3.76	.94	.56	.77	-.4	.65	-.7	1.24	.33	.27	6	6
76	20	3.80	3.76	.94	.56	.77	-.4	.65	-.7	1.24	.33	.27	19	19
76	20	3.80	3.76	.94	.56	.77	-.4	.65	-.7	1.24	.33	.27	25	25
76	20	3.80	3.76	.94	.56	.77	-.4	.65	-.7	1.24	.33	.27	26	26
76	20	3.80	3.76	.94	.56	.77	-.4	.65	-.7	1.24	.33	.27	27	27
75	20	3.75	3.70	.65	.52	.70	-.7	.62	-1.0	1.36	.38	.31	4	4
75	20	3.75	3.70	.65	.52	.70	-.7	.62	-1.0	1.36	.38	.31	7	7
75	20	3.75	3.70	.65	.52	.82	-.3	.77	-.5	1.19	.34	.31	38	38
75	20	3.75	3.70	.65	.52	.82	-.3	.77	-.5	1.19	.34	.31	42	42
74	20	3.70	3.64	.39	.49	.63	-1.0	.58	-1.2	1.48	.43	.34	28	28
73	20	3.65	3.57	.17	.46	.55	-1.4	.53	-1.4	1.60	.47	.37	32	32
72	20	3.60	3.51	-.04	.44	.99	.0	1.00	.1	.92	.36	.40	1	1
72	20	3.60	3.51	-.04	.44	.73	-.6	.73	-.7	1.31	.44	.40	22	22
72	20	3.60	3.51	-.04	.44	.46	-1.7	.47	-1.7	1.72	.52	.40	23	23
72	20	3.60	3.51	-.04	.44	.46	-1.7	.47	-1.7	1.72	.52	.40	24	24
72	20	3.60	3.51	-.04	.44	.46	-1.7	.47	-1.7	1.72	.52	.40	33	33
72	20	3.60	3.51	-.04	.44	.78	-.5	.76	-.6	1.25	.43	.40	41	41
72	20	3.60	3.51	-.04	.44	.78	-.5	.76	-.6	1.25	.43	.40	44	44
70	20	3.50	3.38	-.40	.41	1.59	1.4	1.79	1.8	.15	.32	.45	3	3
70	20	3.50	3.38	-.40	.41	.52	-1.3	.53	-1.3	1.56	.56	.45	14	14
70	20	3.50	3.38	-.40	.41	.52	-1.4	.53	-1.4	1.56	.56	.45	31	31
69	20	3.45	3.32	-.57	.40	2.33	2.5	2.64	3.1	-.75	.25	.48	2	2
69	20	3.45	3.32	-.57	.40	.38	-1.9	.41	-1.8	1.70	.62	.48	12	12
69	20	3.45	3.32	-.57	.40	.39	-1.8	.42	-1.8	1.68	.62	.48	20	20
69	20	3.45	3.32	-.57	.40	.38	-1.9	.41	-1.8	1.70	.63	.48	29	29
69	20	3.45	3.32	-.57	.40	.38	-1.9	.41	-1.8	1.70	.63	.48	30	30
69	20	3.45	3.32	-.57	.40	1.39	1.0	1.30	.8	.47	.41	.48	37	37
68	20	3.40	3.26	-.73	.39	2.08	2.1	2.37	2.6	-.42	.30	.50	9	9
68	20	3.40	3.26	-.73	.39	.25	-2.6	.29	-2.4	1.82	.70	.50	11	11
68	20	3.40	3.26	-.73	.39	.25	-2.6	.29	-2.4	1.82	.70	.50	13	13
68	20	3.40	3.26	-.73	.39	.25	-2.6	.29	-2.4	1.82	.70	.50	21	21
68	20	3.40	3.26	-.73	.39	.25	-2.6	.29	-2.4	1.82	.70	.50	34	34
68	20	3.40	3.26	-.73	.39	.25	-2.6	.29	-2.4	1.82	.70	.50	36	36
68	20	3.40	3.26	-.73	.39	1.51	1.2	1.41	1.0	.31	.42	.50	43	43
67	20	3.35	3.20	-.88	.38	1.85	1.8	2.13	2.3	-.16	.34	.52	10	10
67	20	3.35	3.20	-.88	.38	.40	-1.8	.43	-1.7	1.61	.65	.52	35	35
65	20	3.25	3.08	-1.15	.36	2.14	2.2	2.21	2.4	-.60	.37	.55	16	16
64	20	3.20	3.02	-1.27	.35	.62	-.9	.65	-.8	1.20	.62	.56	5	5
64	20	3.20	3.02	-1.27	.35	3.60	4.1	4.29	4.9	-1.16	.25	.56	8	8
64	20	3.20	3.02	-1.27	.35	3.32	3.8	3.74	4.3	-1.06	.25	.56	17	17
71.3	20.0	3.56	3.46	.00	.48	.96	-.4	1.03	-.3		.43		Mean (Count: 44)	
4.1	.0	.20	.25	.96	.16	.77	1.7	.92	1.8		.17		S.D. (Population)	
4.1	.0	.21	.26	.97	.17	.78	1.7	.94	1.9		.17		S.D. (Sample)	

Model, Populn: RMSE .51 Adj (True) S.D. .82 Separation 1.61 Strata 2.48 Reliability .72
Model, Sample: RMSE .51 Adj (True) S.D. .83 Separation 1.63 Strata 2.51 Reliability .73
Model, Fixed (all same) chi-squared: 116.3 d.f.: 43 significance (probability): .00
Model, Random (normal) chi-squared: 29.4 d.f.: 42 significance (probability): .93

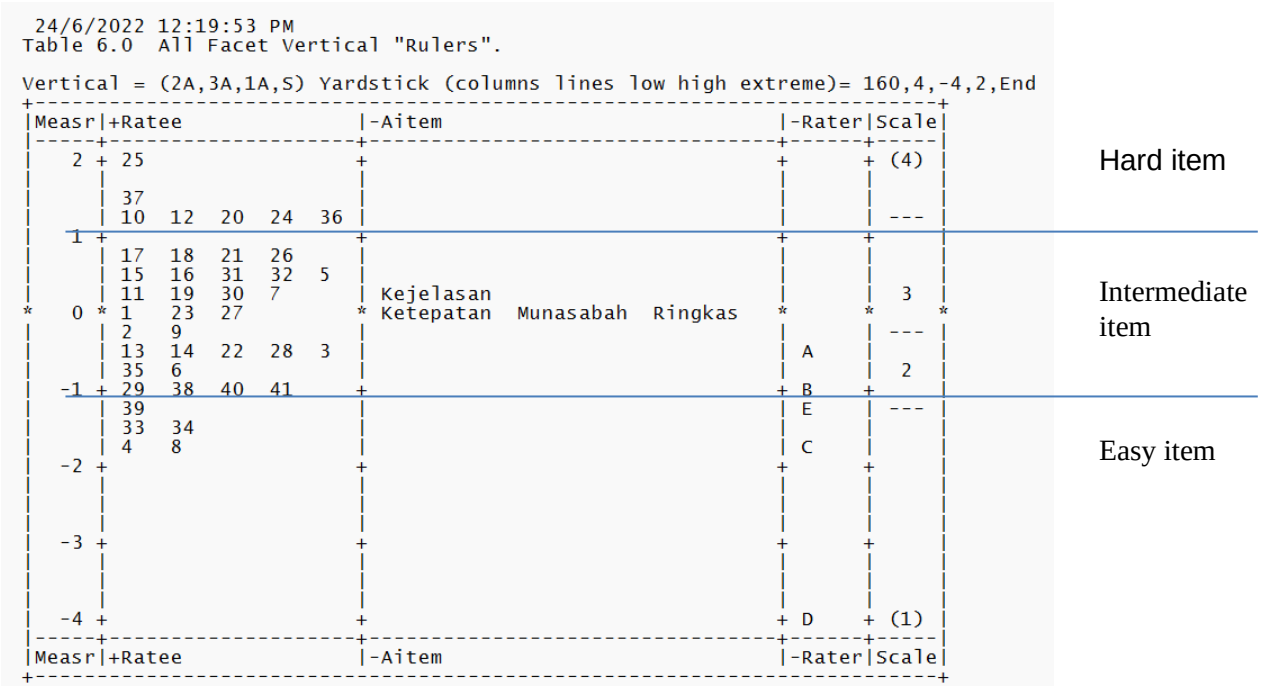
4.3.4 Attitude Construct

Figure 9 illustrates the finding construct map for the construct attitude of item difficulty, item evaluation, and expert panel evaluation levels. The construct map showed a good and acceptable reading of most of the attitude construct items, where the strength of the items was divided into three categories: challenging, moderate, and manageable. Based on the items and expert evaluation, no item was seen below the logit value of -2, which is an item that is not suitable for measuring attitudes. The item evaluation of the attitude item also demonstrated that it measures the whole thing well, except for some items with

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

a higher difficulty level, which were items 25, 37, 10, 12, 20, 24, and 36. The findings for the level of expert evaluation were divided into three categories: medium, easy, and too easy in scoring items. There was a straightforward rater in giving a score, which was Expert A, three accessible Experts, namely B, E, and C, and one expert who was too easy, D. The quality of the analyzed items can be seen in Figure 9, where the higher the item on the logit value, the more both the items used for ISTRI-CBA in the MFRM Measurement Model (George & Stefanie, 2018).

Figure 9: Construct Map for Attitudes



Analysis for statistical compatibility of items for attitude constructs was also determined based on Outfit MnSq (0.5 to 1.5), Outfit Zstd (-2 to +2), or Point Measure Correlation (0.4 to 0.85) values (George & Stefanie, 2018). The overall summary of item quality based on Table 9 found that some items need to be dropped, namely items 6, 8, 27, 34, 37, and 39, as they have Outfit MnSq or Outfit Zstd values that did not match the set values. Appropriate items can measure the same construct, while inappropriate items must be discarded or improved (Villalonga-Olives et al., 2021). The statistical compatibility of the attitude construct items can also be referred to in detail in Figure 10. Based on Table 9, the item separation value was 1.97 less than the value of 2, while the item reliability index was 0.80. The chi-square value was significant, k<0.001, indicating that the items in this attitude construct have good internal consistency. However, some items need to be improved and dropped as those items were considered inappropriate for measuring the attitude construct. Overall, it was found that three items were categorized as very good, 19 items were categorized as good, 13 items could be accepted for improvement, and six items should be dropped.

Table 9: Summary of Statistical Fit Item Categories for Attitudes Construct

Item Category	**Value of *Outfit* MnSq/*Outfit* Zstd**	**Item**	**Total Item**
Very good	1.00	4, 13 dan 28	3

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Good	Between 0.5 to 1.5	5, 10, 11, 12, 18, 20, 21, 22, 23, 24, 25, 29, 30, 31, 32, 35, 36, 38 and 40.	19
Retained/ Modified	<0.5 and >1.5 but the Zstd value is still between -2 and +2	1, 2, 3, 7, 9, 14, 15, 16, 17, 19, 26, 33 and 41	13
Removed/dropped	<0.5 and >1.5 as well as value Zstd <-2 and >+2	6, 8, 27, 34, 37 and 39.	6

Mean = 0.00, S.D = 0.81
Separation = 1.97, Strata = 2.96, Reliability index = 0.80
Chi-square = 239.6, d.f = 40, significant (probability) = 0.00

Figure 10: Analysis Fit Item in MFRM for the Attitude Construct

24/6/2022 12:19:53 PM
Table 7.2.1 Ratee Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu	Ratee
78	20	3.90	3.93	1.95	.75	1.05	.2	.94	.3	.95	.13	.21	25	25
77	20	3.85	3.89	1.48	.63	1.19	.5	4.17	2.1	.59	-.16	.26	37	37
76	20	3.80	3.86	1.14	.55	.99	.1	.79	.0	.99	.24	.29	10	10
76	20	3.80	3.86	1.14	.55	.99	.1	.79	.0	.99	.24	.29	12	12
76	20	3.80	3.86	1.14	.55	.99	.1	.79	.0	.99	.24	.29	20	20
76	20	3.80	3.86	1.14	.55	.99	.1	.79	.0	.99	.24	.29	24	24
76	20	3.80	3.86	1.14	.55	1.07	.2	.93	.1	.91	.17	.29	36	36
75	20	3.75	3.82	.86	.50	.98	.1	.79	.0	.98	.26	.32	18	18
75	20	3.75	3.82	.86	.50	.67	-.5	2.16	1.4	1.03	.29	.32	26	26
74	20	3.70	3.78	.63	.46	1.07	.3	2.14	1.4	.67	.06	.35	17	17
74	20	3.70	3.78	.63	.46	.94	.0	.74	-.1	1.01	.30	.35	21	21
73	20	3.65	3.74	.43	.43	.91	.0	.71	-.2	1.03	.32	.37	5	5
73	20	3.65	3.74	.43	.43	.61	-.7	1.59	.9	1.07	.35	.37	15	15
73	20	3.65	3.74	.43	.43	.57	-.8	1.55	.9	1.12	.38	.37	16	16
73	20	3.65	3.74	.43	.43	.91	.0	.71	-.2	1.03	.32	.37	31	31
73	20	3.65	3.74	.43	.43	.92	.0	.73	-.2	1.01	.30	.37	32	32
72	20	3.60	3.70	.26	.40	.39	-1.5	.37	-1.2	1.49	.64	.39	7	7
72	20	3.60	3.70	.26	.40	.87	-.1	.67	-.4	1.06	.34	.39	11	11
72	20	3.60	3.70	.26	.40	.39	-1.5	.37	-1.2	1.49	.64	.39	19	19
72	20	3.60	3.70	.26	.40	.87	-.1	.67	-.4	1.06	.34	.39	30	30
71	20	3.55	3.66	.10	.38	.33	-1.8	.34	-1.4	1.54	.67	.41	1	1
71	20	3.55	3.66	.10	.38	.46	-1.2	1.19	.5	1.22	.45	.41	23	23
71	20	3.55	3.66	.10	.38	2.52	2.4	3.56	2.9	-1.06	-.13	.41	27	27
69	20	3.45	3.58	-.16	.35	.25	-2.4	.28	-1.7	1.63	.74	.44	9	9
68	20	3.40	3.55	-.28	.33	.59	-1.0	.49	-1.0	.83	.66	.46	2	2
67	20	3.35	3.51	-.38	.32	.49	-1.4	.43	-1.3	.90	.68	.47	3	3
67	20	3.35	3.51	-.38	.32	1.81	1.7	1.66	1.2	.78	.47	.47	14	14
67	20	3.35	3.51	-.38	.32	.45	-1.6	.91	.0	1.24	.45	.47	22	22
66	20	3.30	3.47	-.48	.31	1.47	1.2	1.00	.1	1.17	.57	.49	13	13
66	20	3.30	3.47	-.48	.31	1.47	1.2	1.00	.1	1.16	.57	.49	28	28
64	20	3.20	3.40	-.66	.29	.19	-3.4	.18	-2.8	1.79	.93	.51	6	6
63	20	3.15	3.36	-.75	.29	.97	.0	.72	-.5	1.44	.64	.52	35	35
61	20	3.05	3.28	-.90	.27	.78	-.6	.89	-.1	1.41	.61	.55	40	40
60	20	3.00	3.24	-.98	.27	.78	-.6	.59	-1.1	1.53	.64	.56	29	29
60	20	3.00	3.24	-.98	.27	.52	-1.7	.45	-1.7	1.83	.74	.56	41	41
59	20	2.95	3.20	-1.05	.27	.53	-1.7	.69	-.8	1.65	.68	.57	38	38
55	20	2.75	3.02	-1.32	.25	.30	-3.5	.42	-2.2	1.94	.76	.61	39	39
54	20	2.70	2.97	-1.38	.25	2.24	3.5	2.07	2.7	-.34	.31	.62	34	34
51	20	2.55	2.81	-1.57	.25	1.79	2.5	1.72	2.0	-.07	.41	.64	33	33
50	20	2.50	2.75	-1.63	.25	1.84	2.7	1.77	2.1	.29	.41	.65	8	8
47	20	2.35	2.56	-1.82	.25	1.18	.7	1.02	.1	.01	.62	.68	4	4
68.1	20.0	3.41	3.54	.00	.39	.93	-.2	1.07	.0		.43		Mean (Count: 41)	
8.0	.0	.40	.33	.91	.12	.52	1.5	.81	1.3		.24		S.D. (Population)	
8.1	.0	.40	.34	.92	.12	.53	1.5	.82	1.3		.24		S.D. (Sample)	

Model, Populn: RMSE .41 Adj (True) S.D. .81 Separation 1.97 Strata 2.96 Reliability .80
Model, Sample: RMSE .41 Adj (True) S.D. .82 Separation 2.00 Strata 3.00 Reliability .80
Model, Fixed (all same) chi-squared: 239.6 d.f.: 40 significance (probability): .00
Model, Random (normal) chi-squared: 33.6 d.f.: 39 significance (probability): .71

4.2.5 Professionalism Value Construct

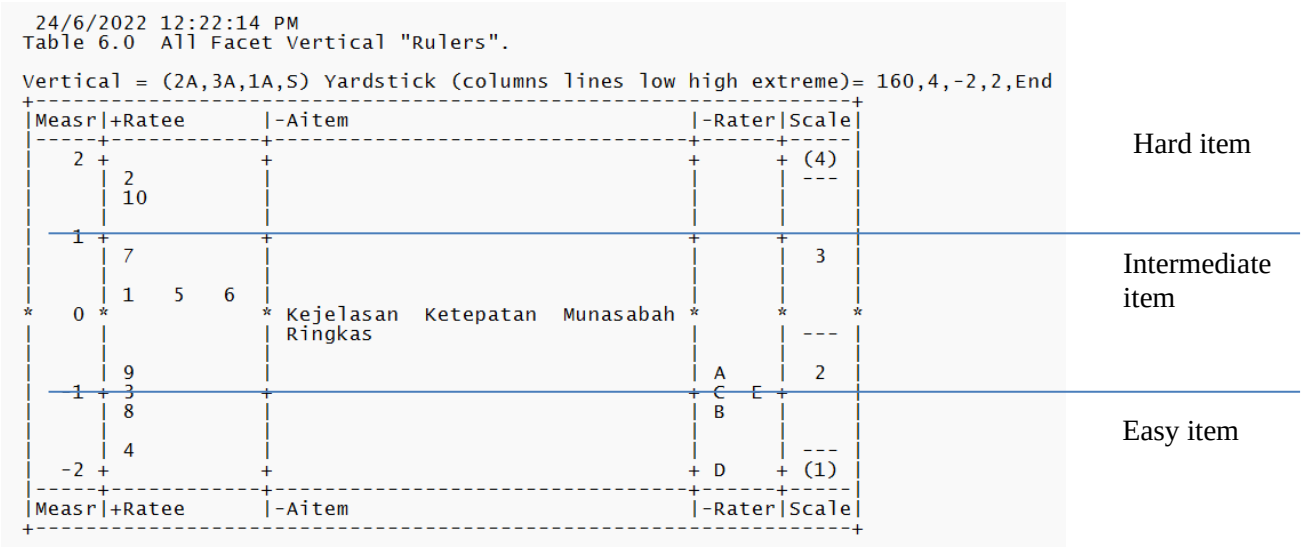
Finally, the findings fit item analysis for the professionalism value construct in Figure 11 were related to item difficulty, item evaluation level, and expert evaluation level, showing a good reading value and acceptable items. This construct map revealed the difficulty level of items divided into three categories: complex, medium, and accessible. Based on the findings of expert evaluation for the items, no item was observed to be below the value of logit -2 of the professionalism value construct. The item-by-item evaluation was also found to measure all except for three items with a higher level of difficulty, namely items 2, 7, and 10. For the expert evaluation level, the items can be divided into two categories: medium for experts A, B, C, and E and easy for expert D to rate items. Figure 11 also shows

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

that the higher the item on the logit value in the construct map, the better the items will be used for ISTRI-CBA in the MFRM Measurement Model (George & Stefanie, 2018).

The MFRM analysis for the following professionalism value construct also looked at the statistical compatibility of items based on Outfit MnSq, Zstd, or Point Measure Correlation values (George & Stefanie, 2018). Table 10 summarizes the compatibility of professionalism value items obtained based on Outfit MnSq and Outfit Zstd not needing to be retained, removed, or dropped. The statistical compatibility of the items of the professionalism value construct can also be referred to in detail in Figure 12, which shows whether the items are to be retained, refined, or discarded.

Figure 11: Construct Map for Professionalism Values



Based on Table 10, the item separation value was 2.58 more than the value of 2.0, while the item reliability index was 0.87. The chi-square value was significant, k<0.001, which suggests that the items contained in the CBA professionalism value have an excellent internal consistency even though there were also improvements for moderate items in the statistical compatibility of the professionalism value construct. Overall, one item was considered excellent, and seven were considered good. Two items were accepted for improvement, and no items were dropped.

Table 10: Summary of Statistical Fit Item Categories for Professionalism Values Construct

Item Category	**Value of *Outfit* MnSq/*Outfit* Zstd**	**Item**	**Total Item**
Very good	1.00	2	1
Good	Between 0.5 to 1.5	10, 7, 5, 6, 3, 8 and 4	7
Retained/ Modified	<0.5 and >1.5 but the Zstd value is still between -2 and +2	1 and 9	2
Removed/dropped	<0.5 and >1.5 as well as value Zstd <-2 and >+2	-	0

Mean = 0.00, S.D = 1.03
Separation = 2.58, Strata =3.78, Reliability index = 0.87
Chi-square = 77.9, d.f = 16, significant (probability) = 0.00

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Figure 12: Analysis Fit Item in MFRM for the Professionalism Values Construct

24/6/2022 12:22:14 PM
Table 7.2.1 Ratee Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu	Ratee
76	20	3.80	3.75	1.73	.55	1.05	.2	1.19	.5	.87	.20	.26	2	2
75	20	3.75	3.69	1.45	.51	.71	-.7	.66	-.8	1.30	.36	.29	10	10
72	20	3.60	3.50	.80	.43	.71	-.7	.74	-.7	1.30	.42	.38	7	7
69	20	3.45	3.31	.30	.39	1.30	.8	1.26	.7	.60	.40	.46	5	5
69	20	3.45	3.31	.30	.39	.51	-1.5	.52	-1.4	1.56	.56	.46	6	6
68	20	3.40	3.25	.15	.38	1.97	2.1	1.91	2.0	.04	.33	.48	1	1
61	20	3.05	2.81	-.70	.33	1.75	1.8	1.84	2.0	.41	.47	.59	9	9
58	20	2.90	2.63	-1.01	.31	.60	-1.3	.58	-1.4	1.32	.71	.63	3	3
55	20	2.75	2.44	-1.29	.30	.77	-.7	.75	-.8	1.46	.70	.66	8	8
50	20	2.50	2.12	-1.73	.30	.60	-1.7	.61	-1.6	2.00	.78	.72	4	4
65.3	20.0	3.27	3.08	.00	.39	1.00	-.2	1.01	-.2		.49		Mean (Count: 10)	
8.4	.0	.42	.52	1.10	.08	.49	1.3	.49	1.3		.18		S.D. (Population)	
8.8	.0	.44	.55	1.16	.09	.52	1.4	.52	1.4		.19		S.D. (Sample)	

Model, Populn: RMSE .40 Adj (True) S.D. 1.03 Separation 2.58 Strata 3.78 Reliability .87
Model, Sample: RMSE .40 Adj (True) S.D. 1.09 Separation 2.74 Strata 3.99 Reliability .88
Model, Fixed (all same) chi-squared: 77.9 d.f.: 9 significance (probability): .00
Model, Random (normal) chi-squared: 8.1 d.f.: 8 significance (probability): .42

4.3 Discussion

The advantage of using MFRM analysis in the validation of ISTRI-CBA items is that it is possible to see the factors of change found in the strictness of raters, which can be controlled through detection, correction, and an unbiased rating scale. The assessment performed simultaneously by field experts on STRI-CBA dimensions allowed the determination of validity and reliability according to the categories of easy, medium, and challenging items. The MFRM approach involves a combination of "expert knowledge and experience" as an aspect of evaluation before being analyzed using Minifac based on items and evaluation criteria (items). The evaluation done by the field experts on the performance of the ISTRI-CBA items revealed a difference in the strictness category of the evaluators. Some experts have the most difficulty and the easiest to give evaluation marks. The difference in evaluation proved that the expert panels assessing the ISTRI-CBA items had various views related to the content of the items. The variety of differences of opinion showed that there was no consensus meeting or discussion between the expert panel before, during, or after the evaluation. Even so, the findings in this MFRM analysis showed that the interaction between the expert panel for all dimensions in ISTRI-CBA was consistent since the findings showed that all the expert panels made a balanced assessment when the percent agreement between the model expectations and the actual value of each construct was very close. Therefore, the findings of the MFRM analysis that determines the internal consistency of the ISTRI-CBA items and the psychometric characteristics of the items were strongly influenced by the effect of the expert panel and the consistency of scoring in expert evaluation.

This study used the item fit index to examine MFRM testing for consistency and the effect of item characteristics on the ISTRI-CBA construct's content validity. The item value matching index was referred to the evaluation of Outfit MnSq or Outfit Zstd. Since the item suitability index reflects the level of measurement of the item that has been developed well or not based on the expert opinion of the appraiser with the conditions set according to the Outfit MnSq or Outfit Zstd stipulations for the item, it was retained, refined or dropped from the inventory. Therefore, the research findings for item suitability for the CBA knowledge construct have a separation index value of 1.52, less than 2, and an item reliability index of 0.70. While the chi-square value was significant, which was k=0.00, the items contained in this CBA knowledge have a moderately good level of internal consistency. The CBA knowledge construct also displayed unidimensionality, with the variance value explained in the Rasch measure being 39.58%, which was very close to 40.0%. The item fit statistics for construct skills of CBA have a separation index of 1.83

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

and an item reliability index of 0.77. The chi-square value was significant, which is k=0.00. These findings showed that the items in this CBA skill have a moderately good level of internal consistency. The CBA skill construct also showed unidimensionality, with the variance value explained by Rasch's measurement being 47.71%. Next, for the third construct, resource support, the statistical fit statistic for the item separation index was 1.61, while the item reliability index was 0.72. The chi-square value was significant, which is k=0.00. This result suggests that the items contained in the resource support have a level of internal consistency that was also moderately good. The resource support construct has a value of variance explained by the Rasch measure of 18.37%, lower than 40.0%. Therefore, to make this reading unidimensional, the items in the source support need to be improved, and inappropriate items need to be removed. The fourth construct, attitude, showed the statistical value of item suitability for the item separation index of 1.97 with the item reliability index of 0.80. The chi-square value was significant, which is k=0.00. These findings showed that the ISTRI-CBA items in the attitude construct have good internal consistency. Next, the unidimensionality characteristic test showed that the value of the variance explained by the Rasch measure was 45.09%. Finally, for the professionalism value construct, the item fit statistic value for the item separation index was 2.58, and the item reliability index was 0.87. The chi-square value was significant as k=0.00, indicating that the items in the professionalism value construct have perfect internal consistency. The variance explained by the Rasch measure in the professionalism value construct was 38.66%, slightly less than 40.0%. The conclusion that can be made from the statistics of the fit of the items against the five constructs is that further improvements need to be made by targeting the items to two separations and unidimensionality by making improvements to the items that were given a warning by the expert panel and dropping items that are not suitable based on Outfit MnSq and Zstd to further increase the validity and the reliability of ISTRI-CBA.

Unique psychometric properties with MFRM analysis can make a practical assessment in bridging the gap between expert panel assessment-related findings with logit values in item fit statistics to ensure the overall validity of the Inventory (John, 1994). The overall findings about the psychometric characteristics of the content validity of the ISTRI-CBA Inventory based on scoring criteria evaluated by an expert panel using MFRM analysis worked very well in ensuring the overall validity and reliability of the instrument developed for further study for perfect testing and improvement.

5. Conclusion

This study presents a new finding on how the STRI-CBA inventory questionnaire items are analyzed and validated. The Rasch Model measurement approach showcased the importance of using MFRM analysis. The MFRM analysis described a good expert panel in evaluating ISTRI-CBA items even from different contents. In addition, MFRM also detailed each item's score to determine the validity obtained from the expert panel of item assessors more reasonably and certify the expert panel of assessors who are qualified and suitable to make assessments based on the analysis results. The consistency of the items in the MFRM analysis can also be shown through the analysis measured by looking at the psychometric characteristics of the items based on the item compatibility statistics and whether the ISTRI-CBA items are retained, improved, or dropped. Therefore, the finding of statistical compatibility allowed the items omitted or improved to measure the readiness of science teachers in CBA on the relevant constructs only. Next, this verified ISTRI-CBA Inventory can be used for further study purposes research, either for pilot studies, validation studies, or field studies to obtain information related to the readiness

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

of science teachers in CBA in primary schools. This study has also proven that four of the five constructs, namely CBA knowledge, CBA skills, attitudes, and values of professionalism, are unidimensional in the ISTRI-CBA Inventory, while the resource support construct demonstrated low unidimensionality. These findings showed dimensions in the resource support construct item to measure the readiness of science teachers in CBA. Therefore, making revisions, improvements, and discarding items that do not fit based on the values of Outfit MnSq and Zstd can improve the unidimensionality reading and the validity of ISTRI-CBA. Overall, the main findings of this study have proven that the ISTRI-CBA developed from the initial 185 items to 162 items have been retained as they have good content validity based on empirical evidence of item analysis based on the MFRM measurement model. Next, this study needs to continue by checking the construct validity and reliability of the ISTRI-CBA Inventory using the Rasch measurement model in a pilot study or a validation study with an actual or larger study sample to further strengthen the level of reliability for the use of ISTRI-CBA in field studies.

Ethics Approval and Consent to Participate

The researchers used the research ethics Educational Research Application System provided by the Research Ethics Committee of Education Planning and Policy Research Division (EPRD), Ministry of Education Malaysia. All procedures performed in this study involving human participants were conducted in accordance with the ethical standards of the ministry research committee.

Acknowledgement

The authors thank the Faculty of Education, Universiti Malaya, Ministry of Education, Malaysia, headmasters and Subject Matter experts involved in this study for good cooperation and support.

Funding

This study received no funding.

Conflict of Interests

The authors reported no conflicts of interest for this work and declared no potential conflict of interest concerning this article's research, authorship, or publication.

References

Azizi Alias & Kamisah Osman. (2018). *Pentaksiran alternatif pembinaan dan pelaksanaan rubrik dalam pendidikan sains.* Penerbit Universiti Kebangsaan Malaysia.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12.

Brown, G. (2013). ASTTLE - a national testing system for formative assessment: How the national testing policy ended up helping schools and teachers. *A developmental and negotiated approach to school and self-evaluation, 14, 39-56. doi:10.1108/S1474-7863(2013)0000014003.*

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Bryant, N. C., & Barnes, L. L. B. (1997). Development and validation of the attitude toward educational measurement inventory. *Educational and Psychological Measurement, 57*(5), *870-875. https://doi.org/10.1177/0013164497057005011.*

Curriculum Development Department. (2019). *Panduan pelaksanaan pentaksiran bilik darjah (2nd ed.)*. Kementerian Pendidikan Malaysia.

Dewan Bahasa & Pustaka. (2017). Kamus dewan edisi keempat-PRPM https://prpm.dbp.gov.my/Cari1?keyword=pentaksiran&d=226112&#LIHATSINI.

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

Examination Board. (2014). Surat siaran Lembaga Peperiksaan Bil. 1 tahun 2014: Penambahbaikan sistem Pentaksiran Berasaskan Sekolah (PBS) Putrajaya: Kementerian Pendidikan Malaysia.

George, E. J., & Stefanie, A. W. (2018). *Invariant measurement with raters and rating scales: Rasch Models for mediated assessments.* Routledge.

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*.

Khairul Azhar Mat Daud, Nik Zulkarnaen Khidzir, Ahmad Rasdan Ismail, & Fadhilahanim Aryani Abdullah. (2018). Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at Pengkalan Datu River. *International Journal of Development and Sustainability, 7*(3), 1026-1037.

Khalil Fakhri Abdul, & Awang Mohd Isha. (2016). Isu Kesediaan Guru dalam Amalan Melaksanakan Pentaksiran Berasaskan Sekolah. *EDUCATUM Journal of Social Sciences, 2*(1), 1-7.

Koloi-Keaikitse, S. (2016). Assessment training: a precondition for teachers' competencies and use of classroom assessment practices [Article]. *International Journal of Training & Development, 20*(2), 107-123. *https://doi.org/10.1111/ijtd.12072.*

Koskimäki, M., Mikkonen, K., Kääriäinen, M., Lähteenmäki, M.-L., Kaunonen, M., Salminen, L., & Koivula, M. (2021). Development and testing of the Educators' Professional Development scale (EduProDe) for the assessment of social and health care educators' continuing professional development [Article]. *Nurse Education Today,* 98. https://doi.org/10.1016/j.nedt.2020.104657.

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.

Madsen, S. R., John, C. R., & Miller, D. (2006). Influential factors in individual readiness for change. *Journal of Business & Management, 12*(2), 93-110. Retrieved from https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=heh&AN=27423392&site=eds-live.

Meisels, S. J. (1998). *Assessing readiness* [Monograph]. Ann Arbor, Mich: [Washington, DC]: Centre for the Improvement of Early Reading Achievement, University of Michigan; U.S. Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Centre, 1998.https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsgpr&AN=edsgpr.000516694&site=eds-live http://catalog.gpo.gov/fdlpdir/locate.jsp?ItemNumber=0466-A-03&SYS=000516694.

Ministry of Education Malaysia. (2017a*). Dokumen standard kurikulum dan pentaksiran KSSM Sains Tahun 3*. WP Putrajaya. Bahagian Pembangunan Kurikulum.

Ministry of Education Malaysia. (2017b). *Dokumen standard kurikulum dan pentaksiran KSSR Sains Tahun 3.* Bahagian Pembangunan Kurikulum.

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Ministry of Education Malaysia. (2018). 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.

Myint Swe, K. (2020). *Rasch Measurement: Applications in Quantitative Educational Research* [Book]. Springer.

Opre, D. (2015). Teachers' Conceptions of Assessment [Article]. *Procedia - Social and Behavioral Sciences,* 209, 229-233. https://doi.org/10.1016/j.sbspro.2015.11.222.

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*.

Rasch, G. (1993). Probabilistic models for some intelligence and attainment tests: ERIC.

Suah, S. L., & Ong, S. L. (2012). Investigating assessment practices of in-service teachers. *International Online Journal of Educational Sciences, 4*(1), 91-106.

Thomas, E. (2015). *Introduction to Many-Facet Rasch Measurement (2nd ed.).* Peter Lang GmbH.

Villalonga-Olives, E., Kawachi, I., & Rodríguez, A. M. (2021). *Rasch model of the bridging social capital questionnaire [Article]. SSM-Population Health, 14. https://doi.org/10.1016/j.ssmph.2021.100791*

Wen-Ta, T., Tzi-Ying, S., & John-Michael, L. N. (2018). Validating translation test items via the Many-Facet Rasch Model. *Psychological Reports*, 1-25. https://doi.org/10.1177/0033294118768664.

Wilson, M. (2005). *Constructing measures: An item response modelling approach*. Lawrence Erlbaum Associates.

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.

Zahari Suppian, & Jamil Ahmad. (2014). Pembinaan dan pengesahan instrumen untuk mengukur tahap kompetensi guru pelatih dalam pentaksiran bilik darjah [Conference publication Non-fiction]. *Seminar Kebangsaan Majlis Dekan-Dekan Pendidikan IPTA, UKM, Bangi.https://ezproxy.um.edu.my/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat00467a&AN=malu.1002517&site=eds-live.*

© 2023 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).