International Journal of Religion
2024
Volume: 5| Number 10 | pp. 256 – 266
ISSN: 2633-352X (Print) | ISSN: 2633-3538 (Online)
ijor.co.uk
DOI: https://doi.org/10.61707/t1ev5t37

Validity and Reliability of the Residency and Immersion Program (PRIme) Assessment Instrument in Enhancing the Quality of New Principals

Aney Marinda Muhammad Amin[1], Norasmah Othman[2] and Aida Hanim A. Hamid[3]

Abstract

The Residency and Immersion Program (PRIme) is an induction for new school principals serving their first year. PRIme was implemented in 2014, and it requires a comprehensive evaluation for program improvement. Hence, this study aims to assess the validity and reliability of the PRIme assessment instrument. The assessment data of content validity by professional experts and face validity by new principals were analyzed using CVR, CVI, and FVI methods. Meanwhile, pilot study data was obtained from 53 respondents and further analyzed using the Rasch measurement model. The questionnaire includes 102 items for assessing the domains of input (courses, coaching, and mentoring), process (type of leadership practiced), and product (leadership and organizational management) in the implementation of PRIme. The Rasch analysis results showed that the items in the questionnaire have good content validity, construct validity, and reliability. Therefore, the Rasch analysis has proven that this questionnaire can be used as an instrument to assess PRIme in an effort to enhance the quality of new principals in Malaysia.

Keywords: *Induction Program, New Principal, Assessment Instrument, Rasch Model, School Leadership*

INTRODUCTION

Induction is one of the key drivers that help new principals develop the necessary skills, confidence, and attitudes for producing high-performing leaders (Rhodes, 2012). It is an important process that orients new principals into their roles and responsibilities, thereby addressing the various issues and challenges faced in schools. By implementing induction through activities such as mentoring and quality guidance, principals who are in new positions can better adapt to new demands in a short period of time (Witten & Marishane, 2021).

The role transition from teacher to leader is a complex learning and reflection process that requires socialization to new practices, identities, and roles (Wardlow, 2008). The socialization process begins as soon as the candidate accepts the new position and faces various task challenges, until the school community finally accepts the newly appointed principal (Lovely, 2004). New principals engage in professional socialization that includes knowledge related to the roles, rules, procedures, processes, and technical skills of a leader. In addition, they are also involved with organizational socialization, which refers to the learning process about the methods and actions that need to be taken after holding the position of principal (Hertting, 2008). The implementation of a comprehensive induction assists new principals in the aspect of professional development through the socialization process more quickly and effectively (Aiken, 2002).

LITERATURE REVIEW

In Malaysia, starting in 2014, all newly appointed school principals are required to undergo an induction program called the Residency and Immersion Program (PRIme). The main goal of PRIme is to ensure that each new principal has a high ability to lead in order to achieve better improvement and enhancement in all aspects of school management (IAB, 2017). New principals completed the program in 13 months through two phases, namely residency (one month), immersion (12 months). The first phase involves new principals to complete a residency program at an identified school with the guidance from experienced principal (mentor). This mentoring program helps them to identify the school culture and critical management issues that can be improved (IAB, 2021). During the second phase, new principals go through professional development courses and coaching sessions regarding strategic and financial management as well as their roles and challenges in

[1] National University of Malaysia, Selangor, Malaysia, E-mail: aneyamin@gmail.com

[2] National University of Malaysia, Selangor, Malaysia, E-mail: lin@ukm.edu.my

[3] National University of Malaysia, Selangor, Malaysia, E-mail: aidahanim@ukm.edu.my

managing the school. These continuous professional learning can lead to improve the knowledge, skills and leadership practices through comprehensive induction program (Joo & Kim, 2016).

The Ministry of Education has implemented various educational development programs from time to time with the hope of changes in the organization, including students, teachers, and administrators, for the better. Such changes refer to the results or effects of program implementation based on the set goals (Yarbrough et al., 2011). However, the effectiveness of a program can be identified after carrying out the assessment, which involves collecting empirical data and detailed information systematically (Chen, 2015). Therefore, the assessment of PRIme should be carried out to ensure that the quality of new principals can be enhanced and reach the set standards. The information obtained can then be used to improve the existing elements of the program and further increase the effectiveness of the program, especially in shaping new high-performing principals. (Gates et al., 2019).

This study evaluates PRIme using the CIPP evaluation model to obtain information to improve the program. This improvement and accountability-oriented evaluation model was chosen after considering its advantages and strengths to comprehensively and systematically evaluate PRIme. PRIme evaluation is conducted by focusing on three levels in the CIPP model, namely input evaluation, process evaluation and product evaluation.

Based on the said requirements, the PRIme assessment questionnaire was developed in reference to the instrument development model presented by Gregory (2015). This model consists of three stages, namely designing, developing, and validating the instrument. The questionnaire was developed based on a review of relevant literature, including the CIPP assessment model (Stufflebeam, 2003), socialization theory (Van Maanen & Schein, 1979), adult learning theory (Merriam & Bierema, 2014), and the second wave of Malaysia Education Quality Standard (JNJK, 2017). Next, the literature material was examined and suited to the research variables and arranged according to the assessment level in the CIPP model, which comprises input assessment, process assessment, and product assessment. Item testing was then conducted through the implementation of a pilot study on new principals who have completed PRIme. The final stage, namely instrument validation, involves the analysis of pilot study data to identify the validity and reliability of the instrument using the Rasch measurement model.

The Rasch model, also known as the One-parameter Logistic model (1PL), applies the Item Response Theory (IRT) to measure item difficulty and individual abilities based on individual responses to the items being tested (Azrilah et al., 2015). This model coordinates data to clearly define measurements (Wisniewski, 1992). It uses the same 'logit' unit to measure item difficulty and individual ability for meaningful comparison (Athanasou & Lamprianou, 2009). The Rasch theory is a robust tool for measuring the development of instruments in the field of Education research, especially in aspects of student achievement and school improvement (Bailes & Nandakumar, 2020; Boone & Scantlebury, 2006). Among the advantages of the Rasch model are (i) the survey is conducted more efficiently, (ii) the time taken for administering the survey is shorter, (iii) the survey data is of high quality, and (iv) the model provides a clear interpretation and direction for the revision of the developed instrument (Bailes & Nandakumar, 2020). In this vein, examination of the instrument through pilot studies is important to ensure that the instrument developed is reliable, valid, and has the appropriate quality to be used for the implementation of the actual study. Therefore, this study was conducted to determine the validity and reliability of the questionnaire in order to assess the implementation of PRIme in enhancing the quality of new principals.

METHODOLOGY

Questionnaire Instrument

Based on the purpose of this study, the instrument was developed to evaluate the implementation of PRIme using the variables proposed in the conceptual framework of the study. This is aimed at ensuring that the developed questionnaire can achieve the objective and answer the research question more accurately and comprehensively. Based on the literature review, items were constructed according to the following variables: (i) input evaluation that includes aspects of course implementation, mentoring, and guidance, (ii) the type of leadership practiced, and (iii) the new principal's leadership and organizational management skills. To collect

data, a five-point Likert scale was used to measure respondents' responses according to the appropriateness of the construct, comprising "Strongly Agree," "Agree," "Slightly Agree," "Disagree," and "Strongly Disagree," as well as "Very Low," "Low," "Moderate," "High," and "Very High." Finally, a PRIme evaluation questionnaire containing 126 items in three sections was developed.

Part 1, which encompasses the assessment of the PRIme implementation input, includes 40 items. Specifically, a total of 13 items were used to measure the implementation aspect of the professional development course, followed by 10 items for measuring the mentoring aspect in Residency and 17 items for measuring the guidance aspect in Immersion. Additionally, the level of respondents' responses to the 40 items was assessed using a five-point Likert scale with the following scores: 1=Strongly Disagree, 2=Disagree, 3=Slightly Agree, 4=Agree, and 5=Strongly Agree.

Part 2 entails an assessment of the PRIme implementation process with 44 items for measuring seven leadership practices among new principals in PRIme. The number of items for strategic leadership is nine, followed by instructional leadership with eight items, cultural leadership with four items, human resource leadership with nine items, managerial leadership with five items, external development leadership with four items, and micropolitical leadership with five items. In addition, the magnitude of leadership skills is assessed on a five-point Likert scale with the following scores: 1=Strongly Disagree, 2=Disagree, 3=Slightly Agree, 4=Agree, and 5=Strongly Agree.

Part 3 of the questionnaire, which encompasses product assessment, consists of 42 items for measuring two variables: leadership skills and organizational management. Both leadership and organizational management variables involve 21 items. Meanwhile, the five-point Likert scale used to measure respondents' feedback is as follows: 1=Very Low, 2=Low, 3=Medium, 4=High, and 5=Very High.

Validity and Reliability

The content validity of the PRIme assessment instrument draft was determined qualitatively and quantitatively by experts in the relevant field. Content validity in this study is based on the Content Validity Ratio (CVR) analysis method and the Content Validity Index (CVI) value (Lawshe, 1975). A total of 12 professional experts assessed each item in the questionnaire using a three-point scale comprising "important," "useful but not important," and "unnecessary." Next, CVR analysis was conducted on the items assessed using the formula $CVR = [ne – (N/2)]/ (N/2)$. CVR values range between -1 and +1, where a value close to -1 indicates that the item is very unnecessary, while a value close to +1 indicates that the item is very important (Lawshe, 1975). The minimum CVR value that must be observed for the 12-expert panel assessment is 0.667 ($N = 12$; $CVR_{critical} = 0.667$) (Ayre & Scally, 2014). Items that do not meet the value must be re-examined to determine whether they should be retained, dropped, or vetted. The CVI value will then be obtained from the entire instrument. Additionally, the minimum CVI value to be observed is not less than 0.78 (Lynn, 1986; Polit et al., 2007). Based on a qualitative review of the items, there were several suggestions put forward by the expert panel for improving the questionnaire. Feedback was also given on items that must be improved in terms of language use and other technical aspects of sentences.

Face validity is another important component in the development of instruments such as questionnaires to support overall validity (Cook & Beckman, 2006). The concept of face validity involves the thought process of the target user of an instrument in supporting the validity of the instrument, and it can be calculated as a Face Validity Index (FVI) (Cook & Beckman, 2006; Yusoff, 2019). In this study, the face validity of the instrument involved 12 new principals who assessed each item using a four-point scale, namely "items are unclear and incomprehensible," "items are unclear and poorly comprehensible," "items are clear and comprehensible," and "items are very clear and very comprehensible." There are two forms of FVI calculation, namely FVI for items (I-FVI) and FVI for scales (S-FVI). Before calculating the FVI, item ratings by the panel for scale 3 or 4 are recoded as 1, while ratings for scale 1 or 2 are coded as 0. Calculation of the I-FVI involves the number of panels that rated the item as 0 divided by the total number of panels. On the other hand, S-FVI is the average score of I-FVI for all items. There are two approaches to calculating S-FVI, namely (i) the average I-FVI score

across all items on the scale (S-FVI/Ave) and (ii) the proportion of items on the scale rated as 1 or universal agreement (UA) across all items (S-FVI/UA).

Population and Sample

The study population involves respondents who had engaged in PRIme across the country. All newly appointed principals are required to take part in PRIme successfully. Therefore, PRIme respondents consist of new principals who were appointed in 2014 and had completed PRIme. The cluster sampling method was used in this study, and it generally involves three levels that encompass clusters by zone, state, and new principal. Meanwhile, the selection of samples for each stage was determined using a simple random sampling technique. This method is typically used to ensure that every subject in the population has an equal chance of being selected as a respondent (Chua, 2006).

The sample size involved in the implementation of the pilot study includes 53 new principals who had participated in PRIme. In terms of the number of respondents, Johanson and Brooks (2010) suggested that 30 representative participants from the target population are a reasonable minimum recommendation for implementing a pilot study for initial survey or questionnaire development. In addition, Linacre (1994) opined that sample sizes as low as 30 to 50 are sufficient to conduct Rasch analysis. Therefore, the selection of pilot study participants, which involves 53 new principals, is sufficient and relevant for the initial survey study or the development of the PRIme assessment questionnaire.

Data Analysis

The validity and reliability of the PRIme assessment questionnaire were determined by analyzing the data obtained from the pilot study using the Rasch measurement model. A total of 53 new principals participated as respondents in the pilot study using a questionnaire comprising 102 items. This takes into account the improvement of items based on the results of content validity assessment by a panel of experts. The items were arranged according to assessment levels encompassing the input (items 1 to 25), process (items 26 to 62), and product assessment (items 63 to 102). Subsequently, the pilot study data was analyzed to identify the items accepted in the questionnaire based on the values of content validity and construct validity, as well as the reliability and separation indices.

The content validity of the research instrument refers to the compatibility of the items. Item compatibility statistics can identify the extent of data compatibility with the Rasch model (Azrilah, 2010). Generally, item compatibility analysis is determined based on the Mean Square (MNSQ) infit value and outfit value. Good item fit has one or both MNSQ values that are in the range of 0.50 to 2.00 (Myford & Mislevy, 1995). In addition, the Z-Standard value (ZSTD) is used to detect whether there is a discrepancy between the data and the model. Accordingly, accepted ZSTD values range from -2.00 to +2.00 (Bond & Fox, 2015). However, if the MNSQ infit and outfit values are accepted, then the ZSTD values can be ignored (Linacre, 2007).

Construct validity can be measured using Principal Component Analysis (PCA) and item polarity. Thus, a principal component analysis was conducted to ensure the dimensional consistency of the instrument (Azrilah et al., 2015). The results of this analysis can improve the construct through item quality compliance (Jusoh et al., 2018). The minimum value to be met for the raw variance explained by the Rasch measurement is 20 percent (Conrad et al., 2012). On the contrary, the value of unexplained variance in the first factor or level of interference is in the range of 5 to 10 percent (Linacre, 2007). Meanwhile, item polarity assumptions are made to determine the ability of items to measure the same construct and the ability of all items to measure a single sub-construct (Bond & Fox, 2007). The polarity value of the item is determined in reference to the Point Measure Correlation (PTMEA CORR) value. Thus, PTMEA CORR values ranging from 0.20 to 0.79 logits indicate that the items can measure the developed construct (Linacre, 2002). However, if the value is positive, i.e., greater than 0, then the item is also accepted (Bond & Fox, 2007).

Analysis of the entire instrument was also carried out based on the reliability index and the item and individual separation index. Individual reliability refers to the consistency of responses for individuals, while item reliability refers to the adequacy of items to measure a construct (Fisher, 2007). Based on the Rasch model approach, the

reliability index refers to the value of internal consistency (Cronbach's Alpha). An acceptable Cronbach's Alpha value is between 0.71 and 0.99, while a reliability value above 0.80 is deemed very good (Bond & Fox, 2015). The item separation index is used to describe the range of item difficulty levels, while the individual separation index is used to describe the range of individual ability levels when answering the questionnaire. Based on the recommendations of Linacre (2018), a value greater than or greater than 2.0 indicates a good separation index.

RESEARCH FINDINGS

CVR and FVI

The content validity process of the questionnaire involved a panel of 12 professional experts to evaluate the PRIme assessment instrument draft consisting of 126 items. The results of item evaluation based on the CVR and CVI methods are shown in Table 1. The study found that 24 items did not reach the CVRcritical value (N = 12; CVRcritical = 0.667). These items include items 4, 5, 6, 12, 17,18, 22, 38, 40, 43, 44, 48, 50, 63, 66, 68, 106, and 107 (CVR = 0.500), items 8, 35, 36, and 39 (CVR = 0.333), and items 26 and 28 (CVR = -0.167). Meanwhile, the CVI value also shows that the 24 items did not meet the minimum value of less than 0.780 (Lynn, 1986; Polit et al., 2007). As a result, the items were dropped from the PRIme assessment questionnaire, and the number of items that remained after the assessment by the expert panel was 102. In addition, the expert panel provided qualitative feedback by suggesting several item improvements in various aspects such as the appropriateness and suitability of items in each construct, the use of language in sentences and statements, as well as the technical aspects and writing format.

Table 1. CVR and CVI value based on ratings of the relevancy of items by 12 experts.

CVR	CVI	Item	Total	Interpretation
1.000	1.000	1, 2, 14, 15, 16, 20, 21, 23, 25, 27, 30, 31, 32, 33, 41, 42, 45, 46, 49, 51, 52, 54, 55, 56, 58, 61, 62, 65, 67, 69, 70, 71, 72, 74, 76, 81, 82, 83, 85, 86, 87, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 101, 102, 103, 104, 105, 108, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 125, 126	69	Appropriate
0.833	0.917	3, 7, 9, 10, 11, 13, 19, 29, 34, 47, 53, 59, 64, 77, 78, 79, 80, 84, 88, 94, 100, 109, 110, 111, 112, 113, 120	27	Appropriate
0.667	0.833	24, 37, 57, 60, 73, 75	6	Appropriate
0.500	0.750	4, 5, 6, 12, 17, 18, 22, 38, 40, 43, 44, 48, 50, 63, 66, 68, 106,107	18	Eliminated
0.333	0.667	8, 35, 36, 39	4	Eliminated
<0.333	<0.417	26, 28	2	Eliminated

The face validity of the instrument involved 12 new principals who were selected as a panel of assessors. The panel conducted assessments for each item. The analysis results for both approaches showed that the I-FVI and S-FVI values were above the minimum value of 0.83 (Marzuki et al., 2018; Yusoff, 2019). Specifically, a total of 91 items reached an I-FVI value of 1.00, while 10 items obtained a value of 0.92 and only one item had a value of 0.83, as shown in Table 2. Meanwhile, the index for universal agreement (S-FVI/UA) was 0.89 and 0.99. From these analysis results, it can be deduced that all items in the PRIme assessment questionnaire have reached a good and acceptable level of face validity.

Table 2. I-FVI value based on the ratings of the items' clarity and comprehensibility by 10 target users

I-FVI	Item	Total	Interpretation
1.000	1, 2, 14, 15, 16, 20, 21, 23, 25, 27, 30, 31, 32, 33, 41, 42, 45, 46, 49, 51, 52, 54, 55, 56, 58, 61, 62, 65, 67, 69, 70, 71, 72, 74, 76, 81, 82, 83, 85, 86, 87, 89, 90, 91, 92, 93, 95, 96, 97, 98, 99, 101, 102, 103, 104, 105, 108, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 125, 126	91	Appropriate
0.833	3, 7, 9, 10, 11, 13, 19, 29, 34, 47, 53, 59, 64, 77, 78, 79, 80, 84, 88, 94, 100, 109, 110, 111, 112, 113, 120	10	Appropriate
0.667	24, 37, 57, 60, 73, 75	1	Appropriate
	S-FVI/Ave. = 0.99		

S-FVI/UA = 0.89

Rasch Model

The analysis results showed that all items have met the MNSQ infit and outfit values from 0.54 to 1.78, as presented in Table 3. Notably, two items, namely items 30 and 33, recorded an infit value of 0.54 and an outfit value of 0.46. However, Myford and Mislevy (1995) opined that good item compatibility has one or both MNSQ values in the range of 0.50 to 2.00. Thus, the two items were retained in the questionnaire. Next, the ZSTD value obtained was between -2.5 and 2.4 for infit and -2.6 and 2.4 for outfit. However, if the MNSQ infit and outfit values are accepted, then the ZSTD values can be ignored (Linacre, 2007).

Based on the results of the principal component analysis in Table 4, the amount of raw variance explained by the measurement is 30.3 percent, which is above the minimum value of 20 percent. Meanwhile, the level of item interference with 5.3 percent is within the set and acceptable range. The raw variance ratio explained by the measure with the variance of the first principal component is 3:11.1, which exceeds the minimum ratio of 3:1 (Conrad et al., 2015). This proves the existence of a unidimensional construct in the developed instrument. Based on item polarity values, all items obtained PTMEA CORR values ranging from 0.19 logits to 0.76 logits (Table 3). The values have met the accepted PTMEA CORR value of between 0.20 logits and 0.79 logits. Although one item obtained a value of 0.19, the value is still positive because it exceeds 0; therefore, the item is also accepted (Bond & Fox, 2007).

The analysis of the entire PRIme assessment instrument was carried out in reference to the reliability index as well as the item and individual separation index as shown in Table 5 and Table 6. The analysis results showed an individual reliability index of 0.96 and an item difficulty reliability index of 0.84. Meanwhile, the Cronbach's Alpha value was 0.97. The findings suggest that the questionnaire items are reliable. Subsequently, the individual separation index was 5.51 and the item separation index was 2.46. These are above the minimum value and are considered very good (Bond & Fox, 2007; Linacre, 2018).

Table 3. Fit statistics of measurement items

Construct	Measure	Model SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD	PTMEA CORR	Item
Courses	-0.08	0.32	0.60	-1.8	0.53	-2.0	0.66	1
	1.02	.30	1.24	0.8	1.33	1.1	0.42	4
	-0.89	0.31	1.33	1.6	1.33	1.4	0.28	5
	-0.79	0.31	1.25	1.2	1.32	1.3	0.44	8
Mentoring	1.86	0.23	0.95	-0.1	1.47	1.4	0.29	9
	1.62	0.25	1.38	1.2	1.19	0.7	0.34	11
	1.56	0.26	1.44	1.3	1.69	1.9	0.36	13
	1.86	0.23	1.19	0.7	1.63	1.8	0.20	14
	1.91	0.23	1.32	1.1	1.62	1.8	0.36	15
Coaching	0.54	0.32	1.38	1.31	1.49	1.5	0.37	17
	0.64	0.31	0.75	-0.8	0.76	-0.7	0.40	18
	0.02	0.32	0.60	-1.8	0.52	-2.0	0.48	23
	0.64	0.31	1.61	1.8	1.56	1.7	0.25	25
Strategic Leadership	-0.79	0.31	1.25	1.2	1.32	1.3	0.44	28
	0.02	0.32	0.58	-1.9	0.52	-1.9	0.64	29
	-0.49	0.32	0.54	-2.5	0.46	-2.6	0.70	30
	-0.18	0.32	1.27	1.1	1.31	1.1	0.43	31
Instructional Leadership	-0.49	0.32	0.54	-2.5	0.46	-2.6	0.69	33
	-1.18	0.31	1.21	1.2	1.18	0.9	0.41	35
	-0.39	0.32	1.58	2.3	1.67	2.3	0.5	36
	-0.79	0.31	0.69	-1.7	0.74	-1.2	0.61	37
Cultural Leadership	-0.39	0.32	1.36	1.5	1.38	1.4	0.54	40
Human Resource Leadership	0.02	0.32	0.81	-0.7	0.77	-0.8	0.58	43
	-0.39	0.32	1.61	2.4	1.72	2.4	0.48	44
	-1.83	0.31	1.14	1.0	1.70	2.6	0.29	46
Managerial	-0.29	0.32	0.57	-2.1	0.47	-2.4	0.63	49

Leadership	-1.08	0.31	1.12	0.7	1.33	1.5	0.43	50
	-0.59	0.32	1.20	1.0	1.30	1.2	0.45	51
	-1.83	0.31	1.31	2.0	1.47	1.9	0.19	52
External	0.02	0.32	1.58	2.0	1.59	1.9	0.37	55
Development	-0.29	0.32	1.54	2.1	1.62	2.1	1.62	57
Leadership								
Micropolitical	-1.65	0.31	0.82	-1.2	0.73	-1.3	0.58	58
Leadership								
Leadership	0.54	0.32	1.31	1.1	1.36	1.2	0.64	63
Skills	-0.18	0.32	0.62	-1.8	0.5	-2.2	0.66	64
	-0.59	0.32	0.65	-1.9	0.57	-2.0	0.72	68
	0.13	0.32	0.84	-0.6	0.72	-1.0	0.65	69
	-0.29	0.32	0.68	-1.5	0.61	-1.6	0.76	71
	-0.29	0.32	1.20	0.9	1.21	0.8	0.68	72
	0.34	0.32	1.71	2.2	1.78	2.2	0.59	73
	-0.08	0.32	0.60	-1.8	0.53	-2.0	0.66	74
	-0.29	0.32	0.68	-1.5	0.63	-1.5	0.65	76
	0.02	0.32	0.62	-1.7	0.56	-1.7	0.60	77
	-1.08	0.31	1.36	1.9	1.36	1.6	0.40	78
	-0.39	0.32	0.77	-1.0	0.74	-1.0	0.60	79
Organizational	0.13	0.32	0.84	-0.5	0.78	-0.7	0.49	83
Management	-0.79	0.31	0.69	-1.7	0.64	-1.7	0.62	84
	0.34	0.32	0.70	-1.1	0.65	-1.3	0.64	86
	1.19	0.29	1.16	0.6	1.31	1.0	0.58	87
	0.93	0.30	1.12	0.5	1.35	1.1	0.38	88
	0.83	0.31	0.55	-1.6	0.59	-1.5	0.43	89
	1.02	0.30	0.74	-0.8	0.77	-0.7	0.47	90
	1.02	0.30	0.76	-0.7	0.67	-1.1	0.54	91
	-.018	0.32	0.77	-1.0	0.77	-0.8	0.40	92
	-0.8	0.32	0.71	-1.2	0.60	-1.6	0.67	93
	1.10	0.29	1.29	0.9	1.38	1.2	0.29	96
	0.13	0.32	1.68	2.3	1.63	1.9	0.46	97
	-0.08	0.32	0.57	-2.0	0.51	-2.1	0.56	98
	0.34	0.32	0.58	-1.7	0.53	-1.8	0.63	99

Table 4. Standardized residual variance (in eigenvalue units)

CVR	**Eigenvalue**	**Empirical**		**Modeled**
Total raw variance in observations	146.4	100.0%		100.0%
Raw variance explained by measures	44.4	30.3%		31.1%
Raw variance explained by persons	20.3	13.9%		14.2%
Raw Variance explained by items	24.1	16.5%		16.9%
Raw unexplained variance (total)	102.0	69.7%	100.0%	68.9%
Unexplained variance in 1st contrast	7.8	5.3%	7.6%	
Unexplained variance in 2nd contrast	7.0	4.8%	6.8%	
Unexplained variance in 3rd contrast	6.0	4.1%	5.9%	
Unexplained variance in 4th contrast	5.1	3.5%	5.0%	

Table 5. Statistical summary for person

	Raw Score	**Count**	**Measure**	**Real SE**	**Infit**		**Outfit**	
					MNSQ	**ZSTD**	**MNSQ**	**ZSTD**
Mean	425.3	102.0	2.56	0.24	1.01	0.1	1.01	0.1

Standard deviation	25.1	0.0	1.23	0.04	0.47	2.1	0.43	2.2
Real RMSE 0.24		True SD 1.20		Separation 5.00		Person reliability 0.96		

Table 6. Statistical summary for item

	Raw Score	Count	Measure	Real SE	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	221.0	53.0	0.00	0.33	1.00	0.0	1.01	0.0
Standard deviation	8.8	0.0	0.82	0.03	0.29	1.2	0.35	1.3
Real RMSE 0.33		True SD 0.76		Separation 2.30		Person reliability 0.84		

DISCUSSIONS

The results of the study have shown that the PRIme assessment questionnaire is a valid and reliable instrument for assessing the implementation of the PRIme induction program in improving the quality of new principals in schools. The content validity of the questionnaire draft was assessed and approved by a panel of experts using CVR and CVI methods; comments for related items to be improved were also given. The CVR method typically measures the importance of the items in each construct domain, while the CVI indicates the simplicity, relevance, and clarity of the items. These methods have helped filter each item empirically to ensure that the retained items truly represent the content of the construct being measured (Almanasreh et al., 2019). In addition, these methods also serve as strong evidence in making a decision to maintain or drop an item in the instrument (Zainal et al., 2020). Therefore, a total of 24 items that did not reach the minimum values of CVR and CVI were dropped from the questionnaire.

Subsequently, the questionnaire was reviewed by a panel selected among new principals for face validity assessment, particularly in terms of item clarity and comprehensibility. In general, face validity using the FVI method helps researchers identify whether there is ambiguity in the instructions and language used for improvement purposes (McDonald et al., 2003). Based on the analysis results, a total of 91 items reached a UA (universal agreement) value of 1.00, which was fully agreed upon by all panel members. Meanwhile, eleven items must be examined and improved based on feedback from the panel, highlighting aspects such as clarity, layout, and presentation in the questionnaire. This shows that the PRIme assessment questionnaire is clear and understandable, especially from the perspective of new principals, who are the main target of this study.

The importance of the Rasch measurement model has been recognized, particularly in assessing the validity and reliability of instruments in survey studies (Golino et al., 2014; Müller et al., 2015; Thompson, 2009). In this study, the validity and reliability of the questionnaire were measured from the aspect of item compatibility, unidimensionality, item polarity, and the reliability and separation index. In terms of item compatibility, the results showed that two items obtained an outfit value less than the minimum value. However, the item is acceptable because the recorded infit value was 0.54, which exceeds the minimum value (Myford & Mislevy, 1995). However, the items must be reviewed and re-examined for improvement. Based on the ZSTD value, 18 items did not reach the specified minimum value. However, Linacre (2007) opined that if the MNSQ infit and outfit values are accepted, then the ZSTD values can be ignored. Therefore, the items were retained in the questionnaire and must be vetted.

Next, the assumption of unidimensionality has been proven using the Residual Principal Component Analysis method. The gross variance explained by the measurement recorded a value of 30.3 percent, which is above the minimum value. Meanwhile, the level of interference measured or unexplained variance in the first contrast suggests that the value of 5.3 percent is adequate and is within the accepted range. Besides, the ratio rate explained by the measure, which is 3:11.1, is above the minimum value set. Adherence to the assumption of unidimensionality is critical in the Rasch measurement model because it can prove that the items in the instrument have a single capacity, i.e., measuring only one pattern (Sumintono & Widhiarso, 2014). In terms of item polarity, all PTMEA CORR values obtained were positive, and the strength of the item's correlation with

the construct is considered acceptable. However, there was one item below the specified range that item requires further checking and vetting. Overall, this shows that all items in the questionnaire are capable of distinguishing between the abilities of the respondents.

Based on the analysis results, the reliability index values for the items and respondents obtained can be considered very good (Bond & Fox, 2015). In addition, the values for the item and respondent separation index as well as Cronbach's Alpha were above the minimum value and are, therefore, considered good and acceptable (Bond & Fox, 2015; Linacre, 2005). The separation index value is also deemed an indication of the instrument's ability to distinguish between the abilities of high and low-performing respondents (Kook & Varni, 2008; Bond & Fox, 2015). Overall, the findings drawn from the analysis suggest that the reliability of the items in the questionnaire is high and this means that the constructed items are stable.

LIMITATIONS

This study has several limitations. The focus of the study is on new principals who were appointed from 2014 until 2021 and had completed PRIme. The rest of the school, including the administrator (Senior Assistant Teacher), teachers, school staff, students, and even principals who no longer hold the position of principal of the school presently were not involved as respondents. The findings of this study are based on the analysis of a pilot study aimed at assessing the validity and reliability of the questionnaire involving only 53 respondents. Therefore, the results cannot be generalized to all new school principals. In addition, this study only includes three assessment levels in the CIPP model, namely input, process, and product. There are only certain aspects involved in each level of assessment based on the research variables of the leadership standards set by the MoE. Therefore, this study does not include aspects other than the domains and constructs being measured even though these have been implemented by the new principals during the implementation of this program.

CONCLUSION

Overall, each item in the PRIme assessment questionnaire has been reviewed and evaluated by a panel of professional experts and a panel of assessors for content validity and face validity. After item improvements were made based on the findings of the CVR, CVI, and FVI analyses, a pilot study was conducted to assess the validity and reliability of the questionnaire using the Rasch measurement model. In summary, the data from each item in the PRIme assessment questionnaire has successfully met the assumptions of the Rasch model. As a result, 102 items were retained after showing good performance in the aspect of item compatibility, item polarity, unidimensionality, and the reliability and separation index. This outcome has added value to the research field, especially in assessing the effectiveness of the implementation of the induction program to the organizational leadership and management of new principals in schools. In this context, the Rasch analysis has also proven that this questionnaire can be used as an instrument for assessing PRIme in an effort to enhance the quality of new principals in Malaysia.

REFERENCES

Aiken, J. A. (2002). The socialization of new principals: Another perspective on principal retention. *Education Leadership Review 3(1),* 32-40. https://eric.ed.gov/?id=EJ659194

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy 15(2),* 214-221. https://doi.org/10.1016/j.sapharm.2018.03.066

Athanasou, J., & Lamprianou, I. (2009). *A Teacher's Guide to Assessment* (2nd ed.). Victoria: Thomson. https://t.ly/sIUgJ

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development 47,* 79-86. https://doi.org/10.1177/0748175613513808

Azrilah, A. A. (2010). *Rasch Measurement Fundamentals: Scale Construct and Measurement Structure.* Kuala Lumpur: Integrated Advance Publishing. https://t.ly/QebAC

Azrilah, A. A., Mohd Saidfudin, M., & Azami, Z. (2015). *Asas Model Pengukuran Rasch: Pembentukan Skala dan Struktur Pengukuran*. Penerbit UKM. https://pnm.overdrive.com/media/3308024

Bailes, L. P., & Nandakumar, R. (2020). Get the most from your survey: An application of Rasch analysis for education leaders. *International Journal of Education Policy & Leadership 16(2),* 1-19. https://doi.org/10.22230/ijepl.2020v16n2a857

Bond T. G., & Fox C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Mahwah, NJ: L. Erlbaum. https://doi.org/10.4324/9781315814698

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410614575

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research, utilizing multiple choice tests. *Science Education, 90(2),* 253-269. https://doi.org/10.1002/sce.20106.

Chen, H. T. (2015). *Practical Program Evaluation: Theory-driven Evaluation and the Integrated Evaluation Perspective* (2nd ed.). Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781071909850

Chua Y. P. (2006). *Research Methods and Statistics Book 2: Statistics Basic*. Kuala Lumpur: Mc Graw Hill Education (Asia). https://t.ly/602yk

Conrad, K. M., Conrad, K. J., Dennis, M. L., & Funk, R. (2012). Validation of the self help improvement scale to the rasch measurement model gain methods report 1.0. Chicago. https://doi.org/10.1177%2F0193841X15599645

Conrad, K. M., Conrad, K. J., Passetti, L. L., Funk, R. R., & Dennis, M. L. (2015). Validation of the full and short-form self-help involvement scale against the Rasch measurement model. *Evaluation Review 39(4),* 395-427. http://dx.doi.org/10.1177/0193841X15599645

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine 119,* 166.e7-166.e16.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions 21(1),* 1095. https://www.rasch.org/rmt/rmt211m.htm

Gates, S., Baird, M., Master, B., & Chavez-Herrerias, E. (2019). *Principal Pipelines: A Feasible, Affordable, and Effective Way for Districts to Improve Schools*. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2666/RAND_RR2666.pdf

Golino, H. F., Gomes, C. M. A., Commons, M. L., & Miller, P. M. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin, 19(3),* 37-54. https://doi.org/10.1037/h0100589

Gregory, R. J. (2015). *Psychological Testing: History, Principles and Applications, Global Edition* (7th ed.). Pearson. https://t.ly/k62-V

Hertting, M. (2008). Are we supporting new principals?. *American School Board Journal 195(6),* 36-37. https://tinyurl.com/6tvt9b66

Institut Aminuddin Baki (IAB). (2017). *Laporan Tahunan 2016 Institut Aminuddin Baki*. Bandar Enstek: IAB, KPM. https://tinyurl.com/548rr99s

Institut Aminuddin Baki (IAB). (2021). *Program Latihan 2020 Institut Aminuddin Baki*. Bandar Enstek: IAB, KPM. https://tinyurl.com/548f5kzr

Jemaah Nazir dan Jaminan Kualiti (JNJK). (2017). *Standard Kualiti Pendidikan Malaysia Gelombang 2*. Putrajaya: JNJK, KPM. https://tinyurl.com/2p88x8kn

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement 70(3),* 394-400.

Joo, H. J., & Kim, T. Y. (2016). Leadership development for school principals: An adult learning perspective. *Journal of Educational Administration and Policy 1(1),* 29-40. http://dx.doi.org/10.22553/keas.2016.1.1.29

Jusoh, M. S., Din, M. S., Hazliza, D., & Suriani, T. (2018). Construct validity for measuring entrepreneurial readiness among Malaysian higher education students: A stochastic measurement model approach. *MATEC Web of Conferences 150,* 06044. https://doi.org/10.1051/matecconf/201815006044

Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the pediatric quality of life inventory 4.0 (PedsQL) generic core scales in school children and adolescents using the Rasch model. *Health and Quality of Life Outcomes, 6, 41*. https://doi.org/10.1186/1477-7525-6-41

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology 28,* 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Linacre J. M. (2018). *A User's Guide to FACETS Rasch-Model Computer Programs Version 3.81.0*. www.winsteps.com.

Linacre, J. M. (1994). *Many-facet Rasch measurement* (2nd ed.). MESA Press. https://tinyurl.com/yn4m3e5t

Linacre, J. M. (2002). *What do Infit and Outfit, MeanSquare and Standardized Mean?*. https://www.rasch.org/rmt/rmt162f.htm

Linacre, J. M. (2005). *Standard Errors: Means, Measures, Origins and Anchor Values*. https://www.rasch.org/rmt/rmt193e.htm

Linacre, J. M. (2007). *A User's Guide to WINDTEPS Rasch-Model Computer Programs*. Chicago, Illinois: MESA Press. https://tinyurl.com/2yw97dp9

Lovely, S. (2004). *Staff the Principalship: Finding, Coaching, and Mentoring School Leaders*. Alexandria, VA: Association for Supervision and Curriculum Division (ASCD). https://archive.org/details/staffingprincipa0000love

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research 35(6),* 381-5. https://www.doi.org/10.1097/00006199-198611000-00017

Marzuki, M. F. M., Yaacob, N. A., & Yaacob, N. M. (2018). Translation, cross-cultural adaptation, and validation of the Malay Version of the System Usability Scale Questionnaire for the Assessment of Mobile Apps. *JMIR Human Factors 5(2),* e10308. https://doi.org/10.2196/10308

McDonald, J. A., Nancy, B., Victor, G. C., & Renee, L. J. (2003). *Questionnaire Design. Reproductive Health Epidemiology Series Module 4*. Atlanta, Georgia. https://stacks.cdc.gov/view/cdc/11674

Merriam, S. B., & Bierema, L. L. (2014). *Adult Learning: Linking Theory and Practice*. San Francisco: Jossey-Bass. https://archive.org/details/adultlearninglin0000merr

Müller, S., Kohlmann, T., & Wilke, T. (2015). Validation of the adherence barriers questionnaire – An instrument for identifying potential risk factors associated with medication-related non-adherence. *BMC Health Services Research, 15(1),* 153. https://doi.org/10.1186/s12913-015-0809-0

Myford, C. M., & Mislevy, R. J. (1995). *Monitoring and improving a portfolio assessment system*. Center for Performance Assessment Research Report, Princeton, NJ: Educational Testing Service. https://eric.ed.gov/?id=ED388725

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health 30(4),* 459–67. https://doi.org/10.1002/nur.20199

Rhodes, C. (2012). Should leadership talent management in schools also include the management of self-belief?. *School Leadership & Management 32(5),* 439-451. https://eric.ed.gov/?id=EJ985921

Stufflebeam, D. L. (2003). *The CIPP model for evaluation*. Paper presented at the Annual Conference on the Oregon Program Evaluators Network (OPEN). Portland, Oregon. https://tinyurl.com/bdfz9ecf

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial* (edisi revisi). Trim Komunikata Publishing House, Cimahi, Indonesia. https://eprints.um.edu.my/11413/

Thompson, N. A. (2009). *Ability Estimation with Item Response Theory*. Assessment Systems Corporation. http://surl.li/twuhf

Van Maanen, J., & Schein, E. H. (1979) Toward of theory of organizational socialization. *Research in Organizational Behavior, 1,* 209-264. https://core.ac.uk/download/pdf/4379594.pdf

Wardlow, R. L. (2008). *Induction and Support for Beginning Principals*. (Unpublished doctoral dissertation). University of California, San Diego. https://escholarship.org/uc/item/10p4z9n7

Wisniewski, D. R. (1992). *Mathematical Models and Measurement*. Rasch Measurement Transactions. https://www.rasch.org/rmt/rmt54f.htm

Witten, A., & Marishane, N. (2021). *Career Pathing for Education Leaders and Managers Through Induction*. Department of Basic Education, Republic of South Africa. https://tinyurl.com/3dcwcr6p

Yarbrough, D. B., Shulha, L. M., Hopson, R. K. & Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.). Thousand Oaks, CA: Sage Publications. https://evaluationstandards.org/program/

Yusoff, M. S. B. (2019). ABC of response process validation and face validity index calculation. *Education in Medicine Journal 11(3),* 55-61. http://dx.doi.org/10.21315/eimj2019.11.3.6

Zainal, M. A., Matore, M. E., Musa, W. N., & Hashim, N. H. (2020). Content Validity of Teacher Innovative Behaviour Measurement Instruments Using Content Validity Ratio (CVR) Method. *Akademika, 90*. https://journalarticle.ukm.my/17320/