International Journal of Evaluation and Research in Education (IJERE)
Vol. 10, No. 1, March 2021, pp. 135~141
ISSN: 2252-8822, DOI: 10.11591/ijere.v10i1.20857

Validating the teaching, learning, and assessment quality of Malaysian ECCE instrument

Hishamuddin Ahmad[1], Nordin Mamat[2], Mazlina Che Mustafa[3], Syahida Iryani Mohd Yusoff[4]
[1,2]Faculty of Human Development, Sultan Idris Education University, Malaysia
[3,4]National Child Development Research Centre, Sultan Idris Education University, Malaysia

Article Info

Article history:

Received Jul 12, 2020
Revised Dec 19, 2020
Accepted Jan 28, 2021

Keywords:

Early childhood care
Education quality
Instrument validation
Polytomous IRT
Teachers

ABSTRACT

Several instruments that measure the teaching, learning, and assessment quality have been developed and published. However, a psychometrically sound instrument to measure teaching, learning, and assessment quality in early childhood care and education that suitable for the Malaysian context needs to be validated. Therefore, this study aimed to validate on teaching, learning, and assessment quality in early childhood care and education instrument, which contains 68 items. The sample comprised 3,498, selected by stratified random sampling from a population of all Malaysian kindergarten teachers. Data were analyzed based on the Polytomous Item Response Theory (IRT) using the Xcalibre software. Samejima's Graded Rating Model (SRGM) was found to be the fit model with the data. Unidimensionality assumption and local independence were tested using the exploratory factor analysis and were fulfilled. The instrument's reliability was overall very good (α=0.966) and the construct validity was also fairly fulfilled with the value of 58.17% total variance explained. Therefore, this instrument is suggested to be used as fairly to measure the quality of Malaysian early childhood care and education among teachers so that appropriate follow-up actions can be implemented towards the betterment of early childhood education quality.

Corresponding Author:

Hishamuddin Ahmad
Faculty of Human Development
Sultan Idris Education University
Tanjong Malim, Perak Darul Ridzuan 35900, Malaysia
Email: hishamuddin.a@fpm.upsi.edu.my

1. INTRODUCTION

Studies have highlighted how teachers play a significant role in both the development of children's learning experiences. This means that teachers perceptions of 'Teaching, Learning, and Assessment Quality in early childhood care & education (ECCE)' will influence their decision on children's learning experiences. The study of Shamsiah, *et al.* [1] state that it is necessary to develop an instrument to gauge, in particular perceptions of the quality of the Malaysian ECCE curriculum.

There are several instruments that measuring teaching, learning, and assessment quality have been developed and published. However, this study is focused on 'Teaching, Learning, and Assessment Quality in ECCE' instrument. The instrument development was based on a Malaysian ECCE. A psychometrically sound instrument to measure teaching, learning and assessment quality in ECCE that suitable for the Malaysian context needs to be validated. Therefore, this study aimed to validate on 'Teaching, Learning, and Assessment Quality in ECCE' instrument for teachers' perceptions in Malaysia. The findings may provide

information to help develop a comprehensive framework on the quality of Malaysian early childhood care & education as a whole [2-4].

Polytomous item response theory (IRT) can be applied to invalidate the instrument [5] as used by Hishamuddin, *et al.* [6]. The polytomous IRT model is generalized from the dichotomous IRT model when more than two categories exist. In other words, the polytomous model is for items that are not scored binarily or wrongly/true (1/0). When some items in the test are scored with more than two response categories such as Likert scale, a polytomous IRT model is required instead of the dichotomous IRT model. The study of Ostini and Nering [7] stated that the simplest and most obvious reason for the development of the polytomous IRT model is the fact that polytomous items are widely used and usually applied in the field of psychological measurement. They also highlighted that all responses in the category or Likert scale could be analyzed using this model of polytomous.

There are several polytomous models that exist. Each model illustrates the extent to which IRT basic philosophy for non-binary data. Each polytomous model determines the behaviour of candidates as their latent trait function (often known as ability). Ordered category polytomous items are items in which the response categories have a clear rank associated with the nature of the study. Likert scale items and partial credit cognitive abilities are examples of polytomous items in the form of ordered categories. In literacy, such items are also known as graded responses. samejima's graded rating model (SGRM) and generalized rating scale model (GRSM) are some examples of graded response model in polytomous IRT model [8-10]. This study aims to validate an instrument which measuring teaching, learning, and assessment quality of Malaysian ECCE using polytomous IRT model. Particularly, the study objectives are to assess the reliability and construct validity of the instrument as well as determining the suitability of the instrument to respondents.

2. RESEARCH METHOD

This is a quantitative study with a survey method. It applies the IRT polytomous model in responding to research questions. Respondents from a population of all ECCE teachers in Malaysia were sampled randomly while the survey is conducted. In the context of IRT, a sample which closely resembles the actual population in terms of numbers is preferred to describe the findings of the study [11]. However, the study of DeMars [12] stated that a minimum sample size of 300 is required for an instrument calibration with a polytomous IRT model. In fact, if the sample size was small or less than 300, the study of Guyer and Thompson [5] explained that the chi-square ($\chi 2$) fit statistics used in a polytomous IRT model would always provide statistically insignificant p values. If such a thing happens, it will certainly provide a meaningless interpretation of the analysis results [13, 14]. The sample size of less than 1,500 is insufficiently small under the demanding data conditions and a complex latent mixture, and a sample size of 2,500 seemed to be adequate. A further increase in the sample size has a positive impact on the accuracy of the estimate, especially in small classes. Therefore, a sample of 3498 ECCE teachers in this study is considered adequate to make a generalization of the population in this study.

2.1. Instrument

Teaching, Learning, and Assessment Quality' instrument is part of 'The Quality of Malaysian Early Childhood Care and Education' instrument which is copyrighted by National Child Development Research Centre (NCDRC) located in Universiti Pendidikan Sultan Idris, Malaysia. It consisting four constructs or factors: 'Child Care Provider's/Teacher's Roles And Interaction With Child', 'Environment, Facilities And Resources', 'Teaching Strategies And Learning Management', and 'Observation And Assessment' [15-18].

This instrument was involved in phases such as planning, construction, testing, and validation, as suggested by Cohen and Swerdlik [19]. Some experts were also involved in all the instrument development phases. The 'Teaching, Learning, and Assessment Quality' instrument contains 68 items which respondents were required to provide responses based on the Likert scale from '1'=Never, '2'=Seldom, '3'=Sometimes, and '4'=Often for each item. The information pertinent to constructs and its items are shown in Table 1.

Table 1. Construct and items assessed in teaching, learning, and assessment quality

Code	Constructs	Number of items	Items
1	Child care provider's/Teacher's roles and interaction with child	25	Item 76 to Item 100
2	Environment, facilities and resources	16	Item 101 to Item 116
3	Teaching strategies and learning management	18	Item 117 to Item 134
4	Observation and assessment	9	Item 135 to Item 143
	Total	68	

3. RESULTS AND DISCUSSION

3.1. Data analysis

Before the data were analyzed with an IRT-based software known as Xcalibre, two assumptions had to be fulfilled. The study of Hishamuddin and Eshah [20] found that the unidimensionality and local independence assumptions should be tested before conducting an IRT-based analysis. As such, the exploratory factor analysis (EFA) was utilized to test the compatibility of unidimensional structures with the data and subsequently testing the local independence of items.

From Figure 1, the first eigenvalue was found much greater than the other eigenvalues. According to Ruscio and Roche [21], steep drop factors or 'elbow' followed by sequence factors is to be inspected. They suggest that, if only there was a drop or dominant 'elbow' or bend in the scree slopes, then the assumption of unidimensionality is satisfied. Therefore, it suggests that a unidimensional model is reasonable for this study data which in line with [20], who suggests that a unidimensional model is reasonable for data if the first eigenvalue was found much greater than the other eigenvalues. The results of Hambleton [22] stated that, when the unidimensionality assumption is met, then the local independence is also obtained. Since the unidimensionality assumption of the latent trait measured in this study is considered reasonable, therefore the assumption of local independence is also accepted.

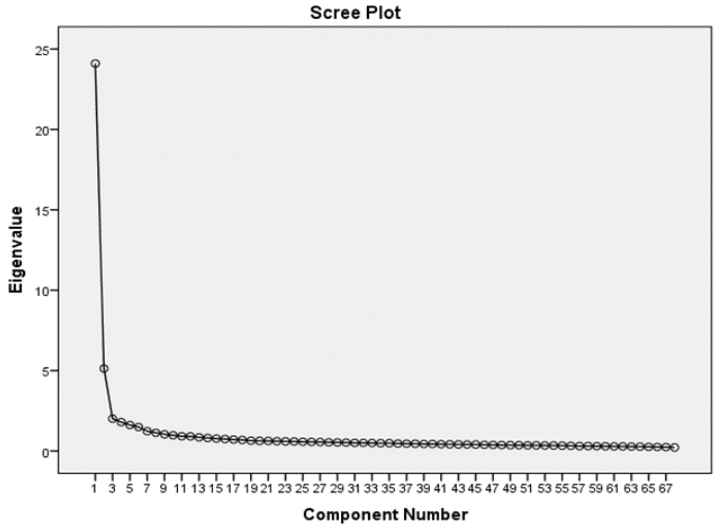

Figure 1. Data scree plot

The Kaiser-Meyer-Olkin (KMO) test (Table 2) with a value of 0.98 indicates that the sample is sufficient for the factor analysis test. The Bartlett's test of sphericity showing chi-square value, $\chi 2$ (2278)=137930.59, $p<0.05$ to identify whether correlations among the 68 items were different to zero, was significant. These indices suggested that the factor analysis test was appropriate and valid to be conducted. Instrument soundness was examined using principal components factor analysis with varimax rotation.

Table 2. Kaiser-Meyer-Olkin measure and Bartlett's test

Kaiser-Meyer-Olkin measure of sampling adequacy		0.98
Bartlett's test of sphericity	Approx. Chi-square	137930.59
	df	2278
	Sig.	0.00

Before the data was analysed with an IRT-based software, two assumptions had to be fulfilled. The study of Hishamuddin and Eshah [20] found that the unidimensionality and local independence assumptions should be tested before conducting an IRT-based analysis. As such, the exploratory factor analysis (EFA) was utilised to test the compatibility of unidimensional structures with the data and subsequently testing the local independence of items. As such, this data analysis was conducted based on the samejima's graded rating model (SGRM) polytomous IRT model for a better fit as proposed by [6, 23].

3.2. Findings

3.2.1. Instrument reliability

In research, the value of α>0.7 is frequently referred to as the 'cut-off value', 'minimum value', or 'good' for reliability index. However, the study of [24] found that the value of α≥0.45 is categorized as 'acceptable' or 'sufficient' to prove the reliability or internal consistency of an instrument. The authors of Griethuijsen, *et al.* [25] in their study to measure students' interest in science in selected countries found a few constructs with α under the value of 0.7 or 0.6. However, this study found that the instrument reliability (α=0.966) is very good and exceeded the minimum value, which was often used as the reference in some researches. Moreover, each of the four constructs is also showing a very good alpha value (Table 3) which means that the reliability of each construct is excellent.

Table 3. Reliability of instrument and factors

Test/Factors	No. of item	Alpha
Full test	68	0.966
1	25	0.959
2	16	0.854
3	18	0.945
4	9	0.877

3.2.2. Instrument construct validity

Validity and reliability are important attributes for the quality of an assessment. The study of Kelley [26] stated that the problem of validity concerns with whether a test really measures what it purports to measure. According to Hair, *et al.* [27], the acceptable total variance explained in factor analysis for a construct to be valid is 60%. However, in the social sciences where information is often less precise, it is not uncommon to consider a solution that accounts for 60% of the total variance (and in some instances even less) as to satisfactory. The authors of Kutluca, *et al.* [28] also suggested that the variance value between 40% and 60% is sufficient for the social sciences. Based on the EFA result of this study, the analysis output showed that the instrument constructs contributed 58.17% of the total variance explained. This indicated that the constructs in the study had sufficient construct validity. As such, it could be stated that the instrument used in this study had fairly measured what it was supposed to measure.

3.2.3. Total information function

The test information function (TIF) is a useful feature of item response theory. It basically tells us how well the test is doing in estimating ability over the whole range of ability scores. The TIF is a graphical representation of how much information the test is providing at each level of theta [5]. According to Hishamuddin, *et al.* [29], IRT begins with the fact that individual responses to items or specific questions are determined by the examinees' mental nature of unobservable or latent traits. In item response theory, the interest is in estimating the value of the ability parameter for an examination. The ability parameter is denoted by θ [30]. IRT allows the latent properties measured on a scale of theta (θ) which has a zero center point in the range from negative infinity to positive infinity. However, the graphs of analysis results with software based on the IRT model, Xcalibre shows a range of θ scale from -4 to 4 [5].

Since a test can be used to estimate the ability of all or each examinee, the amount of information yielded by the test at any ability level can also be obtained [30]. As stated by Baker [30], a test or instrument is a set of items. Therefore, the test information at a given ability level is simply the sum of the item information at that level. Overall maximum total information for full test or instrument used in this study was 101.362 at theta=-2.800 (Table 4). This means the instrument is functioning very well when it is administered to the teachers with less than average on their perceptions for 'teaching, learning, and assessment quality in ECCE' (Figure 2). But, if this instrument is administered to teachers with better perceptions, the TIF will be lower. In other words, teachers' perceptions of the quality of teaching, learning, and assessment quality in early childhood care and education are less than average for this study.

Table 4. Total information function and theta

Test/Construct	TIF	Theta (θ) Level
Full test	101.36	-2.80
1	45.62	-3.00
2	18.30	-2.70
3	31.03	-2.65
4	7.39	-2.45

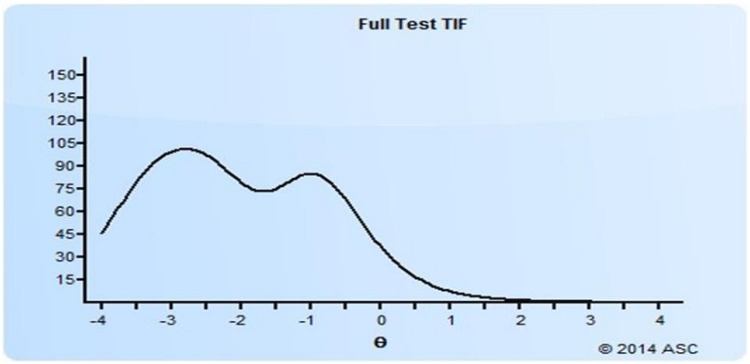

Figure 2. Full test TIF

The situations are similar to all the constructs 1, 2, 3, and 4 (Figure 3). Construct 1 is functioning very well when it is administered to the teachers with less than average (θ=-3.00) on their perceptions for 'child care provider's/teacher's roles and interaction with child'. Construct 2 is also functioning very well when it is administered to the teachers with less than average (θ=-2.70) on their perceptions for 'environment, facilities and resources'. For construct 3, it is functioning very well when it is administered to the teachers with less than average (θ=-2.65) on their perceptions for 'teaching strategies and learning management'. Different from other constructs, construct 4 need the highest theta in functioning very well when it is administered to the teachers. But, the teachers' perceptions for 'observation and assessment' are still less than average (θ=-2.45).

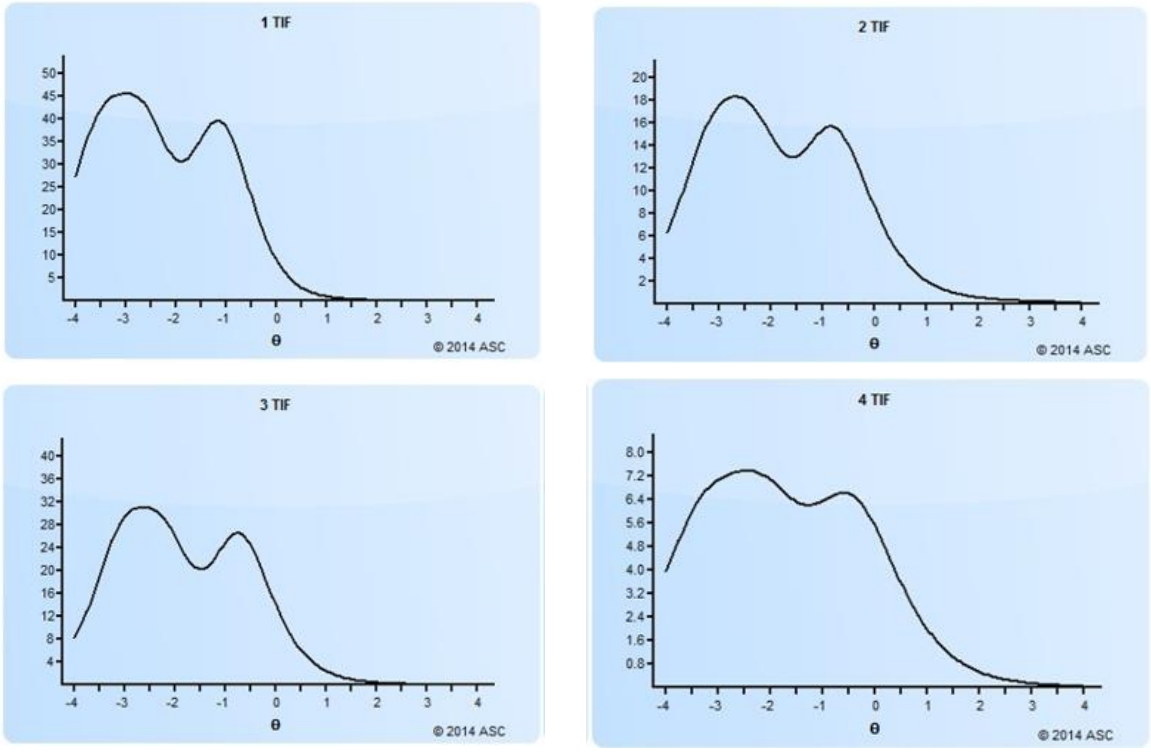

Figure 3. TIF for construct 1, 2, 3 and 4

4. CONCLUSION

Internal consistency for 'teaching, learning, and assessment quality in early childhood care and education' instruments is very high, as stated with α=0.966. The construct validity for the instrument was also found as fairly acceptable, which 58.17% of the items had measured what it was supposed to measure.

Therefore, the instrument is suggested to be used as fairly to measure the quality of Malaysian ECCE among teachers so that appropriate follow-up actions can be implemented towards the betterment of ECCE quality.

At the same time, the instrument is found that it will function better if it is administered to teachers with lower, rather than higher perception to 'teaching, learning, and assessment quality in early childhood care and education'. In other words, this study showed that the teachers' perceptions in 'teaching, learning, and assessment quality in early childhood care and education' is still low. That means is also suggested for the service provider in ECCE to educate and giving well exposures to the teachers about the system as well as 'teaching, learning, and assessment' in their organization. Teachers' perceptions, especially for who are involving in ECCE are very important because the quality in ECCE will also depend on them.

ACKNOWLEDGEMENTS

This part of the research 'Development of a Comprehensive and Integrated Model of Quality Malaysian Early Childhood Care and Education-Project 3/TLA: Determining Contributing Factors of Quality ECCE' (2015-0024-106-04-3) is funded by Ministry of Finance through Ministry of Education Malaysia. Special appreciation is also dedicated to National Child Development Research Centre, Universiti Pendidikan Sultan Idris, Malaysia, for giving us the opportunity to carry out this research project.

REFERENCES

[1] S. Siti Shamsiah, *et al.*, "Parents' Perspectives on Quality ECCE Curriculum: Instrument Development," *International Journal of Engineering & Technology*, vol. 7, no. 4.38, pp. 1266-1269, 2018.
[2] S. Schuck, *et al.*, "The experiences of early career teachers: new initiatives and old problems," *Professional Development in Education*, vol. 44, no. 2, pp. 209-221, 2018.
[3] A. Anis, *et al.*, "Challenges faced by Malaysian private HLIs in providing quality education: a thematic analysis," *Quality Assurance in Education*, vol. 26, no. 2, pp. 349-373, 2018.
[4] L. Heikonen, *et al.*, "Early career teachers' sense of professional agency in the classroom: Associations with turnover intentions and perceived inadequacy in teacher–student interaction," *Asia-Pacific Journal of Teacher Education*, vol. 45, no. 3, pp. 250-266, 2017.
[5] R. Guyer and N. A. Thompson, *User's manual for Xcalibre™ item response theory calibration software, version 4.2*. Woodbury MN: Assessment Systems Corporation Press, 2013.
[6] A. Hishamuddin, *et al.*, "Measuring the academic success of students with ASICS using polytomous item response theory," *International Journal of Advanced and Applied Sciences*, vol. 6, no. 14, pp. 123-129, 2019.
[7] R. Ostini and M. L. Nering, *Polytomous Item Response Theory Models*. Thousand Oaks, California: Sage Publications, Inc, 2006.
[8] N. Mamat, *et al.*, "Reliability, Validity and Fairness of Interviews for Teaching and Learning Project in Quality Early Childhood Care and Education: Preliminary Study," *International Journal of Academic Research in Business and Social Sciences*, vol. 8, no. 6, pp. 1283-1292, 2018.
[9] S. Simon, "A Study of Primary School Parents' Interaction with Teachers' in Malaysia," *World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences*, vol. 11, no. 3, pp. 259-367, 2017.
[10] D. Player, *et al.*, "How principal leadership and person-job fit are associated with teacher mobility and attrition," *Teaching and Teacher Education*, vol. 67, pp. 330-339, 2017.
[11] S. E. Embretson and S. P. Reise, *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc, 2000.
[12] C. DeMars, *Item response theory: Understanding statistic measurement*. Oxford University Press, Inc, 2010.
[13] L. Foong, *et al.*, "Private sector early child care and education in Malaysia: workforce readiness for further education," *Kajian Malaysia*, 36, no. 1, pp. 127-154, 2018.
[14] N. Nordin and M. A. Samsudin, "Professional Identity Formation in Globalization Era: Case of Malaysian Novice Teachers," *Global Business & Management Research*, vol. 9, pp. 193-205, 2017.
[15] C. P. Brown, *et al.*, *The Wiley Handbook of Early Childhood Care and Education*. New Jersey, United States, John Wiley & Sons, 2019.
[16] A. H. Masnan, M. Che Mustafa, and H. Hosshan, "New preschool teachers and implementation of inclusive class's issues in Malaysia," *International Journal of Academic Research in Business and Social Sciences*, vol. 7, no. 10, pp. 644-652, 2017.
[17] C. M. Wotipka, *et al.*, "The worldwide expansion of early childhood care and education, 1985-2010," *American Journal of Education*, vol. 123, no. 2, pp. 1-14, 2017.
[18] A. Church, and A. Bateman, "Methodology and Professional development: Conversation Analytic Role-Play Method (CARM) for early Childhood Education," *Journal of Pragmatics*, vol. 143, pp. 242-254, 2019.
[19] R. J. Cohen and M. E. Swerdlik, *Psychological Testing and Assessment: An Introduction to Test and Measurement*, 5th ed. New York, United States: McGraw-Hill Companies, Inc, 2002.
[20] A. Hishamuddin and M. Siti Eshah, "Is 3PL IRT an appropriate model for dichotomous item analysis of A&P final exam?" *Malaysian Science & Mathematics Education Journal*, vol. 6, no. 1, pp. 13-23, 2016.

[21] J. Ruscio and B. Roche, "Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure," *Psychological Assessment*, vol. 24, no. 2, pp. 282-292, 2012.
[22] R. K. Hambleton, *Fundamentals of item response theory*. Sage Publication, 1991.
[23] R. J. de Ayala, *The Theory and Practice of Item Response Theory*. New York: The Guilford Press, 2009.
[24] S. K. Taber, "The use of Cronbach's alpha when developing & reporting research instruments in science education," *Research in Science Education*, vol. 48, pp. 1273-1296, 2018.
[25] R. V. Griethuijsen, *et al.*, "Global patterns in students' views of science and interest in science," *Research in Science Education*, vol. 45, no. 4, pp. 581-603, 2015.
[26] T. L. Kelley, *Interpretation of Educational Measurements*. New York: World Book Company, 1927.
[27] J. F. Hair, B. J. Babin, R. E. Anderson, and W. C. Black, *Multivariate Data Analysis*, 8th ed. Hampshire, United Kingdom: Cengage, 2018.
[28] T. Kutluca, *et al.*, "Developing a scale to measure information and communication technology utilization levels," *Journal of Turkish Science Education*, vol. 7, no. 4, pp. 37-45, 2010.
[29] A. Hishamuddin, *et al.*, "Detecting item bias in an anatomy & physiology test for nursing students using item response theory," *International Journal of Academic Research in Progressive Education and Development*, vol. 7, no. 1, pp. 97-109, 2018.
[30] F. B. Baker, *The Basic of Item Response Theory*, 2nd ed. Wisconsin: ERIC Clearinghouse on Assessment and Evaluation, 2001.