IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 22, Issue 6, Ver. 3 (June. 2016) PP 04-09
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Rasch Model Implementation inEvaluating Teacher Competency Test Quality on Multimedia Program Expertise

Muhammad Rafie Pawellangi[1*], Wahyu Widhiarso[2], Ahmad Sonhadji[3], Purnomo[3], Hakkun Elmunsyah[3]
[1]*PPPPTK BOE Malang, Indonesia*
[2]*Universitas Gadjah Mada, Indonesia*
[3]*Universitas Negeri Malang, Indonesia*

ABSTRACT:-This study aims to identify and evaluate the psychometric property of Multimedia Program Expertise Competency Test. Research data was obtained from examining 2,744 teachers who teach multimedia subjects in Indonesia. Data were analyzed using Rasch Model to identify psychometric properties at item level (index precision model and point measures correlation) and test level (item-person map, separation index and reliability). Analysis result exhibits that all the items possess high performance as they are in accordance with Rasch Model. At test level, it is also able to accommodate the diversity of teachers' competence level well. On the other hand, a high index of separation obtained exhibits the consistency of items psychometric properties and scores obtained by examined teacher on re-measurement.

Keywords:*RaschModel; Teacher Competency Test; Multimedia Program Expertise*

I. INTRODUCTION

A test should be developed in the best possible quality because its outcome may determine a person's future. Poor quality tests will harm not only test implementer but also individual tested (Anastasi & Urbina, 1997). In poor quality tests, highly competent individuals will gain lower test scores compared to those possessing low competence. As a result, individuals possessing high competence would not be appreciated their ability (Domino & Domino, 2006). They would not be accepted in the selection process, do not benefit economically nor obtain promotional positions to improve their careers. Poor tests are also detrimental to test users as they obtain individuals below-established competency standards which in turn will disrupt human resource management processes.One test that has a considerable impact is Teacher Competency Test (Ujian Kompetensi Guru or UKG). UKG is a very vital aspect for the government to fulfill various teacher management functions in Indonesia. UKG can provide information on teacher competency profile maps nationally as well as assist the government in organizing teacher competency improvement program. Based on these reasons then UKG Test must meet quality measurement standard. Demand for a high quality test is caused by UKG Test’s nature which holds importance related to the quality of teachers nationally.

1.1. Rasch Model Utilization in Evaluating Test Quality

The use of Rasch Model as an approach to examine test quality in Indonesia began to grow rapidly. Researchers in Indonesia are beginning to recognize and understand advantages and usage procedures of analytical techniques introduced by Georg Rasch in the 1960s (Sumintono & Widhiarso, 2013). This approach is very appropriately applied to evaluate UKG Test, considering the importance of this test for the benefit of the nation (Widhiarso, 2016). This section will describe some psychometric properties in Rasch Model perspective that will be used to review and evaluate UKG Test. In general Rasch's psychometric property is divided into two types, i.e property at the item level and test level.

1.2. Rasch Model Property at Item Level

At the item level, properties often reviewed is model accuracy index (fit model) and the correlation between items with Rasch scores (Measures correlation points). The model accuracy index indicates whether item performance has been in accordance with ideal measurement model. Ideal measurement model used is Rasch Model as exhibited in graph relating ability level (X axis) measured by correct answer probability (Y-axis). In exhibited graph, there is a monotonic upward correlation which indicates that higher person's abilities indicate higher probability in answering a problem correctly. The index of model accuracy is indicated by infit or outfit indexes commonly excluded from the Rasch Model-based analysis program. Infit or outfit value between 0.50 and 1.00 are often a reference in exhibiting satisfactory fit model (Bond & Fox, 2007; Meyer, 2014). In addition to indicating item performance suitability and item performance adhering ideal model, the precision value of the model indicates how much the item contribution in measuring target abilities.

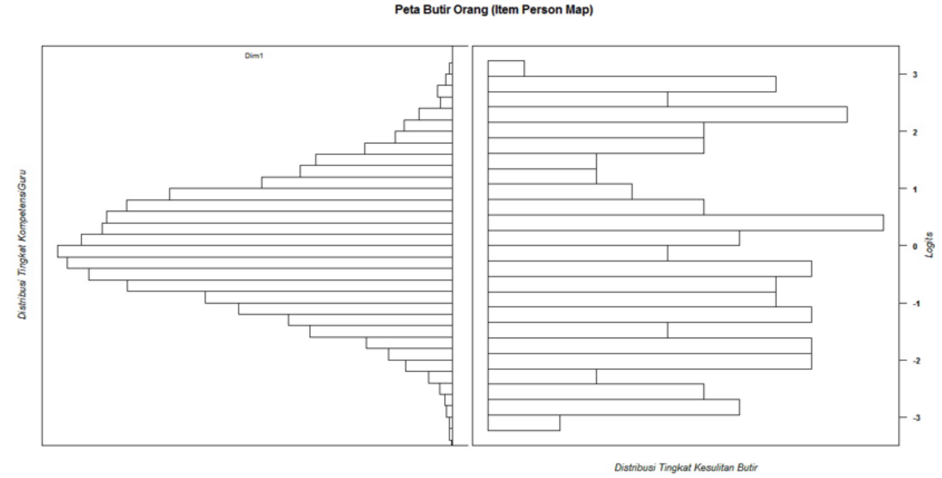


Figure 1. Example of Rasch Analysis Results Item Map

The second property in item level is a correlation between item score and Rasch score. This property is similar to the item-total correlation in the classical test theory approach but the total score on said approach is replaced by Rasch score (Boone, Staver, & Yale, 2014). Rasch score is not a total score obtained by summing correct answer. Rasch Score is obtained from estimating and calibrating analyzed data in the form of logit, therefore its range moves between (approximately) -8 to 8 value. This property exhibits item differentiation.Criteria used are similar tovalue commonly described by measuring instrument development literature, i.e > 0.30.

1.3. Rasch Model Property on Test Level

The psychometric property at test level describes the performance of all items as a single unit. The properties discussed here are separation index and the item-person map. At this level, all items are seen as a single unit of analysis called the test. A separation index is an index exhibiting a disturbance signal(noise) in a data.Separation Index which is one of the mainstay statistics of the Rasch model is obtained from the computational value of the pure score ratio variance divided by the sum of pure score variance and error variance. In the people separation index, the value obtained is equivalent to reliability. A value of 1 indicates a low variance of error and perfect reliability. The value of this index is usually close to the Cronbach alpha coefficient value which is one of the statistics derived from the classical theory approach. For example, a test obtaining a separation index of 0.840, meaning that in the proportion of visible variance score, it contains an 84%pure score variance proportion. According to Boone et al. (2014) separation index ranges from 0 to infinity values,therefore there is no absolute limit for this index. It added that for the purposes of preliminary analysis, higher index values are highly expected compared to low index of separation value. Researchers often use separation index when they conduct experimental research in which the analysis process is carried out on different data. This is due to separation index capability in providing information about data consistencyto be obtained. For example, consistency regarding how consistent the subject score is and how item difficulty consistency level in a test.

In addition to separation index, property at test level is an item-person map or sometimes called a Wright map in memory of the developer, Benjamin Wright. Figure 1 exhibits an example of this map from one of the data analyzed using the Rasch Model. In the figure, individual ability level and item difficulty level are combined in one map as both are converted in the same (metric) unit as logit scale. Logit scale is a common unit of measurement for individual abilities and item difficulty levels. Since both possess the same scale then they can be displayed on a single map. On the other hand, as logit scale is an interval scale of measurement, the distance between one unit and other units of scale is similar. Because of the interval, should an individual named Budi has a logit score of 1 while Agus has a logit score of 2 then it can be concluded that Agus score is twice as much as Budi's. This also applies to item difficulty levels.

An item-person map should exhibit a match between the distribution of individual capability levels and item difficulty levels. In other words, the position of individual ability distribution is in the same location as item difficulty distribution (indicating item difficulty and a person's ability matches on the scale). An item-person map is capable of indicating targeting accuracy that answers whether the test matches the population of the person being tested. The further discrepancy of itemand people position means the lower accuracy of the test to measure individual population tested.

1.4. Overview of Multimedia Program UKG Tests

UKG test for multimedia subjects was developed using teacher's competency model. In this model, teacher competence is divided into two dimensions: pedagogy and professional dimension. These two dimensions are two independent constructs, therefore, treatment of the resulting data needs to be done separately. In other words, they can be viewed as different tests. Analysis result based on UKG tests data exhibit evidence that there are two different dimensions because the correlation between them moves from small to moderate.

1.5. Development ofTeacher Competency Exam onMultimedia Program Expertise

The structure of teacher's competence construct model used as the foundation for this test’s development can be seen in Figure 1. The figure exhibits that competence measurement construct is tiered from competence to item. This model has been tested for accuracy by Pawellangi and Widhiarso (2017) which proves the validity of \competency measurement models constructs used in UKG.

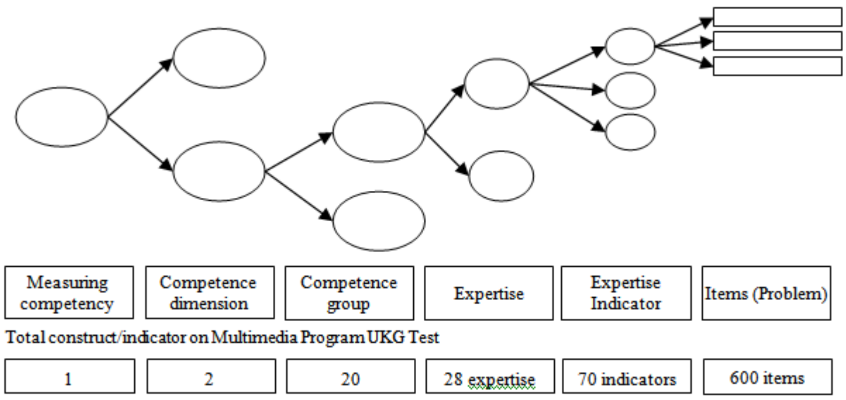


Figure 2. Structure of Teacher Competence Constructive Model

The model described in Figure 2 is a generic model of UKG Test which is generally applied to various subjects tested. In this paper, UKG Tests studied are multimedia subjects tests. Regarding teacher competence model in multimedia subjects, the pedagogic and professional dimensions each contain 10 groups of competencies. Each of these competency groups contains several skill constructs with 18 skills in total. Each skill has several indicators with a total of 70 indicators which are then operationalized into problems. The total problems on the test studied was 600 items.

1.5. Research Objectives

This study aims to test the quality of UKG Multimedia Test both on the item and test level. In addition to identifying these two psychometric property levels, this study also examines the difficulty level of constructs from various levels, from dimensions to items. The research aims to identify and compare what constructs have high difficulty levels that will be a challenge for teachers to overcome. From this, it can be stated that the difficulty level is not only at the item level but also on the higher level of construction. For example at the competence level or competence group dimensions. By exploring existing research results, the comparison of difficulty levels between competency dimensions has not been done. Although pedagogic dimension is seen as a basic skill that most likely mastered by teachers does not mean automatically that difficulty level dimension is lower than professional dimension. This is due to many influential factors, such as subject type as well as teacher’s characteristic.

II. RESEARCH METHOD

2.1. Participants

This research involves vocational school teachers from various provinces in Indonesia. The number of teachers involved was 2,744 people. The highest proportion of teachers was East Java (24.2%), followed by Central Java (19.6%) and West Java (12.6%). Judging from cities or districts distributions where the teachers work, the teachers are spread evenly. Based on teachers’ employment status, the highest proportion is GTY / PTY teachers (30%), School Honorary Teachers (30%) and civil servants (18.9%). Based on teachers’ gender, there are more male teachers (71.6%) than women (28.3%).

2.2. Analysis Procedure

Rasch model-based analysis was undertaken using WINSTEPS 3.2 program (Linacre, 2000). The data analyzed is data taken fromvocational school teachers who took UKG test on multimedia expertise. The properties studied include model accuracy index (fit model) exhibitn by the infit-outfit value, item discriminating power indicated by the correlation index of logit (point of measures), separation index exhibiting the reliability of the measurement and item-person map. The following are the criteria used to decide whether the item or test being examined has expected performance. (A) Items with high precision value are items possessing high accuracy value are infit or outfit index value of 0.5 to 1.5 (Meyer, 2014), (b) item correlation index with a total score above 0,10 or positive, (c) measurements reliability indicated by reliability coefficients of KR-20 above 0.70, and (d) separation index for both items and persons above 2.00 (Cheng et al., 2011). If these criteria can be met then the tests evaluated in this study could be concluded to have fully satisfying psychometric properties.

III. FINDINGS

3.1. Psychometric Property at Item Level

Complete test item properties can be seen in Table 1. Itexhibited that the average difficulty level of items (in logit scale) has a mean of 0.00 with a standard deviation of 1.29. The average value of 0.00 is a determination made by Rasch Model which is performed on all tests where the value is likened to the midpoint of a ruler. This is consistent with occurrence in everyday life where rulers are sedentary but varies in quantity (e.g length) on measured object.

Table 1. Psychometric Properties of Multimedia Expertise UKG Test

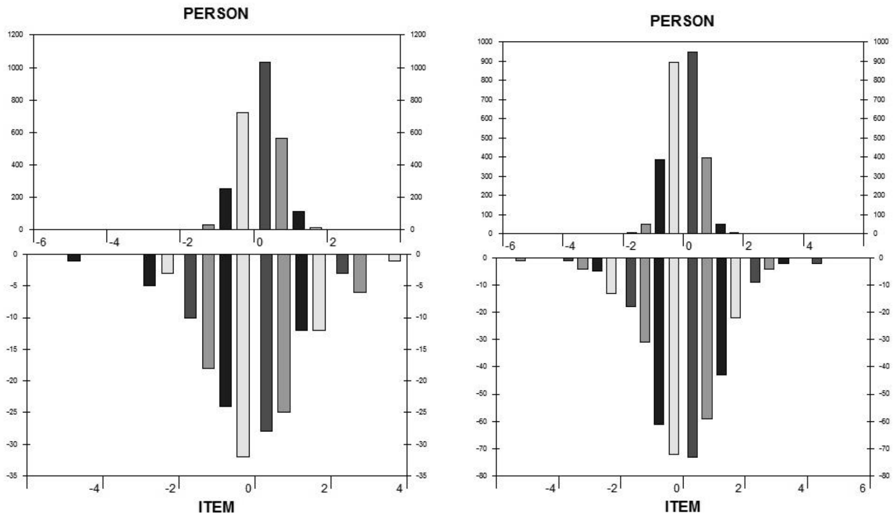
Competency Dimension, Item Properties, Mean, SD, Range
Professional
(Total item 420), Difficulty Level, 0,00, 1,29, -4,22 – 3,60
, Item discrimination, 0,29, 0,04, 0,08 – 0,36
, Infit, 0,99, 0,01, 0,97 – 1,03
, Outfit, 0,99, 0,02, 0,91 – 1,09
Pedagogic
(Total item 180), Difficulty Level, 0,00, 1,26, -4.54 – 3.50
, Item discrimination, 0, 215, 0,040, 0,071 – 0,286
, Infit, 0,999, 0,008, 0,975 – 1,029
, Outfit, 0,999, 0,019, 0,902 – 1,074

In the professional dimension, the range of item difficulty levels (b = -4.22 to b = 3.60) indicates that measured test range is broad enough. The extent of this range indicates that it has an optimal function in measuring, although it is used in individuals with varying competence levels, from very low to very high competencies. Based onitem discrimination point of view, there are no negative items. It indicates that every item contributes in measuring teacher competence. There are some items that have low performance because power value difference is below 0.20. This condition can still be tolerated because there are some low difficulty level and high difficulty level items. In this test, low-power (<0.20) items are either very low or very high difficulty, not caused by miskeying, underutilization or underperformance. The precision value exhibitn by the infit-outfit index indicates that all items have a satisfactory accuracy index because no item has an infit-outfit value below 0.50 or above 1.50. Similar result was also found in test items on pedagogic dimension.

3.2. Psychometric Property at Level Test

The measurement reliability by UKG Tests exhibitn in classic test approach through KR-20 coefficients in both dimensions is quite high, 0.95 for professional dimension and 0.89 for pedagogic dimension. These results indicate that internal consistency of the components in the test (in this case items) is quite high. In other words, the items in the test are homogeneous and possess equivalent measuring precision (Chernyshenko, Stark, Drasgow, & Roberts, 2007).The item separation index in both dimensions possesses adequately high value, 24.62 in professional dimension and 25.3 in pedagogic dimension. The high value indicates the consistency of psychometric properties in test items. Consistency is indicated by no change in item difficulty level or item accuracy index when the test is imposed on teachers other than those involved in the study (Fox & Jones, 1998). In addition to item separation index, Rasch's Model also identifies teachers' consistency if tested again with a similar test. This information is indicated by a person separation index. Person separation index in both dimensions also possessesadequately high value, 4.32 in professional dimension and 2.80 in pedagogic dimension. The high value indicates teacher scores consistency if subjected to tests at other times. On the other hand, measurements reliability on both tests is 0.98 which indicates the high internal consistency of the two subtests.

Figure 3 exhibits the distribution of teacher competence compared to item difficulty level in the two test dimensions. The figure exhibits that item difficulty distribution level has the accuracy on tested teacher competency level. In other words, it can be said that item difficulty distribution levels reach all levels of teacher competence, from the lowest to the highest.



,
Pedagogic Competency, Professional Competency

Figure 3. Results of Item-Person Map in Two Subtests of Teacher Competency Test

3.3. Comparison of Test Indicators Difficulty Levels

Overall, there are four items with the highest degree of difficulty in this. Points with a high difficulty level were 3.31 followed by 2.9, and 2.84. The items are in Device Development Module, Learning Theory and Principles, and Remedial and Enrichment Programs. The three items are said to have the highest level of difficulty because few of the examinees are able to answer these points correctly. While the simplest items are -3.46, -2.77, and -2.59 respectively. These items are Learning Assessment and Evaluation Module, Device Development, ICT Application in Study. These items are said to have a low level of difficulty as most participants are able to answer the item correctly.At competency group level in pedagogical competence, the 9th competency group (Remedial and Enrichment Program) has many items with the highest difficulty level. On the other hand, the fourth competency group (Curriculum Development) has many items with a low degree of difficulty. At indicator level, "Able to explain the features contained in the learning model" and A-10-01-02 indicators (presented the PBM illustration from the Beginning to the end, and "Capable of exhibitingcorrect reflection activities according to the illustration" is an indicator With the highest average difficulty level of 0.98 and 0.87, while the indicator with the easiest level of difficulty is "Ability to explain media concept and/or learning resources as one of the means of achieving the learning objectives as a whole" and "Presented data on process assessment and learning outcomes assessmentof a student, PU is able to determine a Student Report Score based on the data "with an average difficulty level of -0.84 and -0.90.

IV. DISCUSSION

This study aims to evaluate the quality of teacher competence test on multimedia expertise program by using Rasch Model. Item analysis and test models utilizing Rasch Model have not been widely applied in Indonesia. Research on item analysis and test is still dominated by the use of classical test theory (CTT) based analysis. In this study, the items of teacher competency tests related to multimedia expertise were analyzed by identifying psychometric properties at the item level and test level. The results exhibit test item evaluated produced satisfactory properties that are visible from all the items in the test, satisfying high model accuracy to Rasch Model. This high accuracy indicates that all the items in the test run in accordance with its function of measuring attributes.At a higher test level, evaluated test also performs well because the difficulty level of these test items is spread over a variety of subjects and includes a broad continuum of competence levels, from low to high competencies. In other words, this test can fulfill its function as a test to identify the varied competence of teachers in Indonesia. The reliability of the items indicated by the high item separation index values indicatesitem psychometric property consistency despite the fact the test are assigned to different teachers. On the other hand, people reliability exhibiting score consistency obtained by teachers is also high. These results indicate that a teacher measured twice or more will get an equivalent score. The equivalent in terms of scores is also equivalent in relative positions with other teachers. For example, teacher A who has a lower score compared to teacher B, if both are subjected to the same test the results do not change their position, teacher A's score remains under teacher B.These findings can be predicted in advance because the writing of the items on this test went through a fairly lengthy process ranging from reviews of item content relevance by external experts and has gone through a series of item improvement processes ranging from content measurements and grammar improvement. These results indicate that a careful review process will produce items with good performance. The results at the test level corresponding to expectations are due to the writing of items adjusting to the created grid. In said grid, the competence is separated to several domains and indicators as an operational representation of the competence constructs measured. On the other hand, in each section of the grid, there is a division of domains and indicators into several strata or levels, from easy, moderate to difficult levels. This procedure is one of the recommended activities in the development of test kits (AERA, APA, & NCME, 1999). It is this division that allows the items in the test to have a diverse range of difficulty levels, ranging from highly precise items in an endeavor to measure teachers with low and high competencies.

Although the results of the evaluation on this test obtainedis satisfactory, there are still some issues that need to be explored to complement the findings. For example, test validation from both predictive and construct. Questions about this validity will answer how much this test can predict teacher performance in conducting education and teaching in the classroom or how successful the teacher is in fostering his career in education. The question of construct validity will answer whether the constructs measured by this test represent adequate competent constructs and relevant field condition (Cohen & Swerdlik, 2009). In this activity, the analysis conducted is factor analysis capable of identifying kinds of competence construct structure measured by this test. In general, based on the findings of this study, other test developers who want to measure competency levels can take information from the process undertaken in this study, which is the development of teacher's competence test on Multimedia ProgramExpertise. The study aimed for teachers toobtain optimal results as well as results following after undertaking this test. The ideal item-person map for this test can be followed up by developing and administering this test using a computer-adaptive testing (CAT) -based adaptation strategy which is the latest approach in measuring competence.

REFERENCES

[1] AERA, APA, & NCME. (1999). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

[2] Anastasi, A., & Urbina, S. (1997). Psychological Testing. Upper Saddle River, NJ): Prentice Hall.

[3] Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model. Fundamental Measurement in the Human Sciences. Mahwah NJ: Lawrence Erlbaum.

[4] Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch Analysis in the Human Sciences Heidelberg: Springer.

[5] Cheng, K. K. F., Lee, J., Leung, S. F., Liang, R. H. S., Tai, J. W. M., Yeung, R. M. W., & Thompson, D. R. (2011). Use of Rasch Analysis in the Evaluation of the Oropharyngeal Mucositis Quality of Life Scale. Nursing Research, 60(4), 256-263. doi: 10.1097/NNR.0b013e318221f731

[6] Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: toward increasing the flexibility of personality measures. Psychol Assess, 19(1), 88-106. doi: 2007-03014-008 [pii]10.1037/1040-3590.19.1.88

[7] Cohen, R. J., & Swerdlik, M. E. (2009). Psychological testing and assessment: An introduction to tests and measurement. Boston, MA: McGraw−Hill Companies, Inc.

[8] Domino, G., & Domino, M. L. (2006). Psychological testing an introduction. Cambridge: Cambridge University Press.

[9] Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. Journal of Counseling Psychology, 45(1), 30-45. doi: 10.1037/0022-0167.45.1.30

[10] Linacre, J. M. (2000). WINSTEPS, version 3.02. Chicago: Winstep.com.

[11] Meyer, J. P. (2014). Applied measurement with jMetrik. New York, NY: Taylor & Francis.

[12] Pawellangi, M. R., & Widhiarso, W. (2017). Pengujian model kompetensi guru dengan menggunakan analisis faktor konfirmatori. Manuskrip Publikasi. Jakarta: Kementrian Pendidikan dan Kebudayaan.

[13] Sumintono, B., & Widhiarso, W. (2013). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial. Cimahi: Trikom Publishing House.

[14] Widhiarso, W. (2016). Penerapan Model Rasch untuk mengevaluasi Tes UKKS dan UKPS. Tenaga Kependidikan, 1, 50-51.