INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 12, Issue 3, (2023) E-ISSN: 2226-6348

# Validity and Reliability of The Teacher Competency Instrument (TECOMS) for Small Schools in Malaysia: Rasch Model Analysis

Mohd Norlizam Mohd Razali, Aida Hanim A. Hamid, Bity Salwana Alias

Center for the Study of Leadership & Educational Policy, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia

**To Link this Article:** http://dx.doi.org/10.6007/IJARPED/v12-i3/18564 DOI:10.6007/IJARPED/v12-i3/18564

***Published Online:*** 22 September 2023

**Abstract**

This study aims to validate and evaluate the reliability of the Teacher Competency Instrument (TECOMS) to identify the level of teacher competence in small schools. A total of 10 expert panels were involved in validating this instrument. For that purpose, a questionnaire was distributed to 75 teachers who teach in small schools as respondents in this study. A version of the winsteps software version 5.2.2.0 was used to check the items' reliability and item-respondent isolation, the polarity and suitability of the items measuring constructs, and the standardized residual correlations. It also allows the removal or purification of items based on item polarity, item suitability, and determining values based on standardized residual correlation values. The final analysis found that no items were omitted. The final instrument showed a total of 39 appropriate items used to measure teacher competence constructs. The analysis revealed that these items have a good level of validity and reliability and are suitable for use by teachers in small schools.

**Keywords:** Validity and Realibility, Teacher Competency, Instrument, Small School, Rasch Model.

## Introduction

A small school refers to the number of students 150 and below (Ministry of Education, 2012). The number of small schools in Malaysia in 2021 is 30.75% of all schools in Malaysia and the locality of these small schools is 73% in rural areas (Ministry of Education, 2021). Academic achievement in these small schools is on average in the low category when compared to larger schools, the average composite score achievement for small schools is only 68%, and the achievement is 4% behind other schools (Ministry of Education, 2012). One of the factors mentioned is the difficulty of finding and retaining teachers and placing quality head teachers to serve in the school concerned.

Various factors contribute to the achievement of this small school, the majority of which are in rural areas. According to a study conducted by Emil and Sedat (2022) and Mazdi et al (2014) found that there are several factors that can influence student achievement, particularly contribute to poor achievement, among them are socioeconomic factors and student background, school location, teacher's teaching style in the classroom and leadership practices of leaders school that affects student performance in school. MOE has implemented various approaches to ensure that student achievement at every level can be improved, especially in small schools. Through the Malaysia Education Blueprint (MEB) 2013-2025, small schools are also given focus, this concentration is not only from the aspect of school development but also the aspect of improving the achievement of student performance to be comparable to the achievements of other schools. Based on the MEB 2013-2025 in chapter four emphasizing student learning, this chapter also specifically explains MOE's efforts in improving student achievement in small schools through the placement of quality teachers in these schools (Ministry of Education, 2012).

MOE always pays attention to the quality of teachers. This may be observed in the establishment of the Malaysia Teacher Standard (MTS), which was established in 2009 as a reference to the professional competencies of teachers that educators in Malaysia must attain in order to ensure quality teachers and quality service across the service. Apart from that, the Ministry of Education also provides teacher performance evaluation documents in the form of the Integrated Assessment of Education Service Officers, which was published in 2017, and teacher quality standard documents in the form of the Malaysian Education Quality Standard and Malaysian Education Quality Standard Wave 2. These documents produced by the MOE are a guide to determining the quality of teachers in Malaysia. However, the changes in education today lead to the need for changes in teacher competencies when in 2020 the MOE has issued the Malaysian Teacher Standards 2.0 (MTS 2.0) framework, the essence of the framework, concerned set four main domains in the competence namely i) Knowledge Orientation, ii) Instructional, iii) Community Involvement and iv) Personal Quality (Ministry of Education, 2020).

Based on the teacher competency requirements introduced by MOE in the framework of MTS 2.0, there are additional competencies that teachers need to possess, namely the community aspect. These competency requirements in line with MEB 2013-2025, in the ninth shift, have clarified the involvement of parents, community representatives, and private organizations in efforts to upgrade education in schools (Ministry of Education, 2012). This gives a positive indication of the development of students in schools through the community and this study will also look at the community aspect as a subdomain in teacher competence as found in the framework of MTS 2.0. Therefore, the aspect of competence is a factor that contributes to student achievement and the quality of education.

**Problem Statement**

The focus of the Ministry of Education (MOE) Malaysia on small schools through The Malaysia Education Blueprint (MEB) 2013-2025 is to improve performance in a small schools. Based on information from MOE, the achievement in a small school in 2019 showed that the average composite score was below 65% (MOE 2019), and besides that, a statement from the Minister of Education, Science, and Technology Research of Sarawak, Datuk Seri Michael Manyin Jawong, explained that the analysis carried out by the state government found that the cause of low achievement in Malaysian Certificate of Education and Primary School Achievement Test 2014-2016 is due to a small school. Through the report by the Board of Inspectors, there are 17 small school schools that did not get any bands, and 693 small schools got ratings between bands 4 to 7 which belong to the low category (Ministry of Education, 2012). This information describes the low level of academic achievement in small schools. Many studies conducted, including Haslinda et al (2017); Losius et al (2018); Normiati and Abdul Said (2019), show that student achievement problems are due to the learning approach implemented by teachers in schools, and student achievement is related to teacher quality.

Shift 1 in the MEB 2013-2015 which is providing equal access to international quality education is MOE's approach to bridge the achievement gap between urban and rural schools. MOE's efforts to ensure that all schools have equal access to achieve excellent results. MOE has also introduced initiative 56 which aims to improve achievements in small schools because academic achievement in small schools is at a low level, i.e. 33% get bands 4–6 and 2% have no bands (Ministry of Education, 2012). Furthermore, the small school roadmap under initiative #40 in 2017 is a continuation of initiative plan #56 under MEB 2013-2025 to address small school issues from the following aspects: 1) cost-effectiveness in small schools and 2) student performance improvement in a small school (Ministry of Education, 2019). Referring to MOE Professional Circular Letter No. 9/2017 related to the implementation of combined classes with multi-grade teaching and learning in schools with fewer students with an enrollment of 30 and below (MOE, 2017). In a study conducted by Kritis (2016), the implementation of blended classes presents great challenges to teachers, especially when teaching involving students of different ages in the same class. Studies also show that teachers are less confident and skilled in implementing Teaching and learning activities in blended classrooms. Nitce (2020) agrees with this notion, stating that teachers are less confident in conducting blended classes.

Subsequently, the need for teacher posts in small schools was reviewed based on the needs and suitability of school operations to ensure the optimization of the use of human resources in small schools (Ministry of Education, 2017). In 2018, the MOE implemented the abolition of rehabilitation teacher positions on an exchange basis in small schools based on operating expenditure estimates for 2018, which involves a total of 1,043 teachers. This measure involves the posting of rehabilitation teachers in small schools with an enrollment of 100 or below in Peninsular Malaysia and small schools with an enrollment of 50 or below in Sabah and Sarawak (Ministry of Education, 2019). Based on the actions that have been implemented, this has to some extent affected the number of teachers in small schools in general. The workload of teachers in small schools will increase, especially in the operational tasks in schools and teaching that do not follow the options due to the limited number of teachers. Based on the findings of Sharma et al (2021) a person will be stressed if the task is too heavy and they are unable to complete the task and at the same time affect the employee's performance. This situation, if continued, can cause a person to become irritable, act aggressively, and lose interest and concentration in work. According to Carroll et al. (2021), if teachers fail to control themselves due to stress in performing tasks, it will affect students and thus affect the quality of school education.

Based on the above statements, there is a gap in small schools, especially in terms of teacher quality. In order to improve teacher quality, it is necessary to analyze all aspects of competence in detail in order to ascertain the level of teacher competence in small schools. Therefore, researchers have developed the TECOMS instrument among teachers in small schools in Malaysia. A pilot study was conducted to ensure that the questionnaire instrument has good validity and reliability.

**Purpose of The Study**

The purpose of this study is to create a TECOMS instrument for assessing teacher competency in small schools and to thoroughly evaluate the instrument's validity and reliability using the following diagnoses: Dependability and item-respondent separation;

1) Detect the polarity of the item;
2) The fit of the item measures the construct;
3) Determine the dependent items based on the standardized residual correlation values;
4) Examine the dimension uniformity;
5) Item and Respondent Mapping; and
6) Evaluate the best rating scale's validity.

## Methodology

**Research Design**

This study is a cross-sectional study conducted. This study was conducted in two states namely Perak and Negeri Sembilan. This research has obtained ethical approval from the Malaysian Ministry of Education (Reference: KPM.600-3/2/3-eras(11992)), Perak Education Department (Reference: JPNPk.SPS.USJK.600-1 Jld.3(20), and Negeri Sembilan Education Department (Reference: JPNS.SPS.PSR.500-12/4 Jld.5. This research has obtained ethical approval by the Universiti Kebangsaan Malaysia.

**Research Sample/Participants**

The population for this study involves primary schools with the status of small schools in State Perak and State Nine. The selection of the study sample was simple random involving teachers who teach in small schools in the two states that have been mentioned. The selection of the population in these two states is due to the large number of small schools and the variety of types of small schools found in the states. For this study, a total of 150 questionnaires were distributed to 150 teachers in 73 small schools in Perak and Negeri Sembilan only 78 questionnaires were correctly returned and then analyzed.

**Data Collection Method/Instrumentation**

The study conducted uses a questionnaire instrument and is quantitative in nature. The instrument consists of two parts. Part A contains the background of the respondents, such as gender, service background, and school background. Section B contains questions related to teacher competence. A total of 39 items were developed based on the Malaysian Teacher Standard 2.0 (Ministry of Education, 2020) framework. Before the instrument was distributed, it went through a process of face and content validation together with the appointed experts. The expert panel was asked to review and provide views and suggestions on the format, content, and language style structure. A total of five experts with a qualification of Doctor of Philosophy Degree in the fields of Competence, Measurement & Evaluation, and Malay Language with more than five years of experience were appointed as an expert panel to test the level of validity of the instrument.

To determine the inter-expert reliability achieved, the Content Validation Index (CVI) was used. The CVI value accepted or agreed upon among experts is 0.80 and above for new instruments. (Davis, 1992). There are also views suggesting 0.78 and above for three experts and the rest (Polit & Beck, 2006). However, in this study, the researchers adopted a CVI value of 0.8 by Davis (1992) for the new instrument, in line with the intention of the study.

**Data Analysis Method**

In this study, researchers will use the Rasch Measurement Model, A version of the winsteps software version 5.2.2.0 which is able to provide an opportunity for research to prove the validity of the instrument not only based on Cronbach's alpha but also through proof of the quality of the items studied to further strengthen the validity of the instrument (Azrilah et al., 2013).

**Findings and Discussion**

The implemented studies will be assessed using winsteps software utilizing the Rasch measurement model technique. The researcher performed a functional examination of the items from the aspect of (i) reliability and classification of the respondent's items, (ii) detecting the polarity of the items that measure the construct based on the PTMEA CORR value (iii) detecting the appropriateness of the item (item fit), (iv) determining the dependent item based on the residual correlation value standardized, (v) measuring unidimensionality using the Residual Principal Component Analysis (PCA), and (vi) the validity of the best rating scale. The use of this diagnosis is contingent upon the study's objectives, namely the selection and filtering of high-quality things from the assessed items. Items that do not fit the analysis's criteria will be repaired or discarded. The following provides an explanation for each item's functional check.

**Reliability and Item-Respondent Isolation**

Before an instrument can be used in a study, it must undergo a procedure of determining the validity and reliability of the items to assure the instrument's and data's quality. This level is critical to ensuring that the data collected during the actual data research is accurate and capable of answering all of the researcher's research questions.

Typically, the researcher will refer to Cronbach's alpha reliability coefficient to determine the instrument's level of dependability. This dependability rating is used to reference a widely used measurement model based on the True Score Test Theory (TSTT), often referred to as the classical model. However, apart from Cronbach’s alpha value, the Rasch measurement model is based on Item Response Additionally, the theory was discovered to be capable of providing clues to the proof of item quality, thereby enhancing the validity of an instrument developed as a consequence of considerations that take into account the individual's (person's) level of ability and the item's level of difficulty (Azrilah et al., 2013; Rosseni et al., 2009). Cronbach's Alpha values of 0.71–0.99 are considered acceptable when using the Rasch measurement model, as defined by (Bond & Fox, 2015).

Table 1
*Interpretation of Alpha-Cronbach Scores (Bond & Fox, 2007)*

| Alpha-Cronbach Score | Reliability |
|---|---|
| 0.9 - 1.0 | Very good and effective with a high degree of consistency |
| 0.7 - 0.8 | |
| 0.6 - 0.7 | Good and Acceptable |
| < 0.6 | Acceptable |
| < 0.5 | Items need to be fixed |

To determine the reliability of items in the instrument, analysis and statistics with the Rasch measurement model approach were used with reference to the reliability values as well as item segregation. According to the results of the pilot research analysis, the reliability value derived using Cronbach's Alpha (α) value is 0.91, as shown in Table 2. This rating clearly indicates that the instrument is in excellent condition, is effective, and has a high level of consistency, and may be utilised in research.

Table 2
*Reliability Values (Cronbach's Alpha) for the Pilot Study*

```
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .97
```

Following that, the instrument is analyzed holistically by examining the reliability index and isolating items and respondents. According to Table 3, the item reliability is 0.93, while the item isolation is 3.54. This shows that the item reliability index is quite good and effective, approaching the value of 1.0 with a high degree of consistency. The anticipation of repetition is equally strong for this construct when delivered to other groups of respondents with comparable abilities (Bond & Fox, 2015). With a value of 3.77, the isolation index is more than 2.0 (Bond & Fox, 2015), indicating that these items are statistically classified into three strata or levels of agreement.

While Table 4 displays the respondents' reliability and quality value as 0.98, the respondents' isolation value is 4.67. This shows that the responders' reliability is quite strong and satisfactory. According to Bond and Fox (2015), dependability values of more than 0.8 are considered to be satisfactory. Likewise, the respondent's separation value shows a very good separation of the level of difficulty for each item.

Table 3
*Item Reliability Values*

SUMMARY OF 25 MEASURED ITEM

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 252.0 | 75.0 | .00 | .13 | .96 | -.35 | .98 | -.27 |
| SEM | 6.1 | .0 | .09 | .00 | .06 | .36 | .07 | .35 |
| P.SD | 30.0 | .0 | .46 | .01 | .29 | 1.77 | .35 | 1.74 |
| S.SD | 30.6 | .0 | .47 | .01 | .30 | 1.81 | .36 | 1.77 |
| MAX. | 313.0 | 75.0 | .91 | .14 | 1.52 | 2.77 | 1.79 | 3.12 |
| MIN. | 196.0 | 75.0 | -.99 | .12 | .60 | -2.91 | .55 | -2.87 |

```
REAL RMSE   .13 TRUE SD   .44  SEPARATION  3.34  ITEM  RELIABILITY  .92
MODEL RMSE  .13 TRUE SD   .44  SEPARATION  3.54  ITEM  RELIABILITY  .93
S.E. OF ITEM MEAN = .09

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Table 4
Respondent Reliability Values

TABLE 3.1 Kompetensi.xlsx ZOU774WS.TXT Mar 31 2022 11: 0
INPUT: 75 PERSON 25 ITEM REPORTED: 75 PERSON 25 ITEM 5 CATS MINISTEP 5.2.2.0

SUMMARY OF 75 MEASURED PERSON

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 84.0 | 25.0 | .50 | .26 | .95 | -.13 | .98 | -.05 |
| SEM | 3.1 | .0 | .15 | .01 | .03 | .11 | .03 | .11 |
| P.SD | 26.9 | .0 | 1.31 | .10 | .23 | .92 | .30 | .98 |
| S.SD | 27.1 | .0 | 1.32 | .10 | .23 | .92 | .30 | .99 |
| MAX. | 121.0 | 25.0 | 3.08 | .53 | 1.62 | 2.19 | 1.67 | 1.92 |
| MIN. | 42.0 | 25.0 | -1.28 | .17 | .40 | -3.27 | .41 | -2.94 |

REAL RMSE .28 TRUE SD 1.28 SEPARATION 4.55 PERSON RELIABILITY .95
MODEL RMSE .27 TRUE SD 1.28 SEPARATION 4.67 PERSON RELIABILITY .96
S.E. OF PERSON MEAN = .15

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .97

**Detect Item Polarity**

Item polarity analysis (Point Measure Correlation) or item parallelism is an indicator used to show the items used moving in one direction by the measured construct. A measure with a positive index for all things shows that all items are operating parallel to each other. If, however, a negative index exists, the researcher must re-examine the data to determine whether it should be adjusted or discarded.

The first step in assessing the validity of a construct is to examine the item's polarity. The Point Measure Correlation (PTMEA CORR) assessment is used to determine an item's polarity in order to ascertain whether the construction accomplishes or sets the desired goal. According to Bond and Fox (2007), a positive value (+) in the PTMEA CORR section shows that the item measures the construct being measured. If the displayed value is negative (-), the generated item has no effect on the measurement construct. As shown in Table 5, all items had a positive Point Measure Correlation (PMC) value. This indicates no conflict between the item and the construct being measured. However, the data indicate that the positive items go in the same direction as the construct, are capable of measuring the construct, and do not contradict the construct being tested. If the value of PTMEA CORR is high, then the item can differentiate the abilities between respondents.

**Tracking Fit Item Suitability**

The adequacy of the components in the measuring construct can be determined by the MNSQ outfit and infit values. Bond and Fox (2007) recommend that the MNSQ's outfit and infit values between 0.6 and 1.4 to guarantee that the items created are adequate for measuring constructs. However, the MNSQ outfit index needs to be paid attention to instead of infit to determine the fit of items that measure a construct or latent variable (Mohd Kashfi, 2011). If the MNSQ value is more than 1.4 logit, then it means that the item is misleading. If the MNSQ result is less than 0.6 logit, it suggests that the respondent anticipates the item too easily (Linacre, 2007). Additionally, the ZSTD and infit values of ZSTD should be between 2 and +2 Bond & Fox (2007), however, if the MNSQ outfit and infit values are approved, the ZSTD index can be omitted (Linacre, 2007). Based on table 5, it was found that there are four items that are not in the set range and they need to be refined or dropped. Items above the value of 1.40 in the MNSQ outfit column are items 4 (1.79), 1 (1.74) and 6 (1.63), while items below the value below 0.6 are only item 21 (0.55). As a result of this analysis, all four elements will be revised while none will be eliminated, taking into account the needs of researchers and expert opinions.

Table 5
*Item Polarity and Item Fit Suitability*

TABLE 26.1 Kompetensi.xlsx ZOU774WS.TXT Mar 31 2022 11: 0
INPUT: 75 PERSON 25 ITEM REPORTED: 75 PERSON 25 ITEM 5 CATS MINISTEP 5.2.2.0

PERSON: REAL SEP.: 4.55 REL.: .95 ... ITEM: REAL SEP.: 3.34 REL.: .92

ITEM STATISTICS: CORRELATION ORDER

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | JMLE MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD | PTMEASUR-AL CORR. | PTMEASUR-AL EXP. | EXACT MATCH OBS% | EXACT MATCH EXP% | ITEM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 313 | 75 | -.99 | .14 | 1.40 | 1.80 | 1.79 | 2.83 | .40 | .58 | 49.3 | 57.4 | D4 |
| 1 | 290 | 75 | -.57 | .13 | 1.52 | 2.77 | 1.74 | 3.12 | .45 | .64 | 30.7 | 51.6 | D1 |
| 6 | 294 | 75 | -.64 | .13 | 1.46 | 2.41 | 1.63 | 2.67 | .49 | .63 | 33.3 | 52.1 | D6 |
| 16 | 281 | 75 | -.43 | .12 | .91 | -.50 | 1.27 | 1.41 | .58 | .67 | 38.7 | 49.4 | D16 |
| 3 | 277 | 75 | -.37 | .12 | 1.19 | 1.21 | 1.19 | 1.03 | .59 | .68 | 29.3 | 48.8 | D3 |
| 5 | 293 | 75 | -.62 | .13 | 1.09 | .56 | 1.13 | .69 | .61 | .64 | 45.3 | 51.9 | D5 |
| 11 | 280 | 75 | -.42 | .12 | .90 | -.60 | 1.09 | .54 | .62 | .67 | 38.7 | 49.2 | D11 |
| 2 | 273 | 75 | -.31 | .12 | 1.24 | 1.50 | 1.15 | .87 | .63 | .69 | 38.7 | 46.8 | D2 |
| 15 | 263 | 75 | -.16 | .12 | 1.18 | 1.15 | 1.24 | 1.34 | .65 | .71 | 32.0 | 43.8 | D15 |
| 24 | 268 | 75 | -.24 | .12 | 1.43 | 2.51 | 1.16 | .91 | .67 | .70 | 45.3 | 45.4 | D24 |
| 10 | 251 | 75 | .01 | .12 | 1.05 | .37 | 1.07 | .47 | .71 | .74 | 41.3 | 41.3 | D10 |
| 8 | 241 | 75 | .16 | .12 | 1.05 | .37 | .97 | -.12 | .76 | .76 | 26.7 | 38.5 | D8 |
| 7 | 238 | 75 | .20 | .12 | .74 | -1.71 | .84 | -.90 | .77 | .77 | 33.3 | 38.3 | D7 |
| 14 | 252 | 75 | .00 | .12 | .73 | -1.84 | .79 | -1.25 | .79 | .74 | 48.0 | 41.5 | D14 |
| 13 | 212 | 75 | .62 | .13 | 1.17 | .96 | .90 | -.42 | .81 | .82 | 45.3 | 42.7 | D13 |
| 22 | 260 | 75 | -.12 | .12 | .61 | -2.91 | .60 | -2.64 | .82 | .72 | 53.3 | 43.2 | D22 |
| 17 | 243 | 75 | .13 | .12 | .72 | -1.95 | .74 | -1.57 | .82 | .76 | 48.0 | 38.6 | D17 |
| 23 | 227 | 75 | .37 | .12 | .74 | -1.70 | .70 | -1.70 | .85 | .79 | 37.3 | 36.9 | D23 |
| 25 | 235 | 75 | .25 | .12 | .82 | -1.14 | .75 | -1.49 | .85 | .78 | 41.3 | 38.1 | D25 |
| 12 | 222 | 75 | .45 | .13 | .69 | -2.00 | .73 | -1.47 | .87 | .80 | 36.0 | 39.3 | D12 |
| 19 | 232 | 75 | .29 | .12 | .60 | -2.89 | .63 | -2.28 | .87 | .78 | 41.3 | 37.9 | D19 |
| 18 | 211 | 75 | .63 | .13 | .93 | -.31 | .72 | -1.40 | .87 | .82 | 52.0 | 43.3 | D18 |
| 20 | 229 | 75 | .34 | .12 | .60 | -2.86 | .55 | -2.87 | .88 | .79 | 52.0 | 37.9 | D20 |
| 9 | 218 | 75 | .52 | .13 | .70 | -1.92 | .60 | -2.31 | .88 | .81 | 49.3 | 39.2 | D9 |
| 21 | 196 | 75 | .91 | .14 | .63 | -2.06 | .55 | -2.21 | .91 | .84 | 56.0 | 51.4 | D21 |
| MEAN | 252.0 | 75.0 | .00 | .13 | .96 | -.35 | .98 | -.27 | | | 41.7 | 44.2 | |
| P.SD | 30.0 | .0 | .46 | .01 | .29 | 1.77 | .35 | 1.74 | | | 8.0 | 5.7 | |

*Mengukur Keseragaman Dimensi*

Aspek keseragaman dimensi merupakan aspek yang kritikal dalam menentukan sesuatu instrumen itu dapat mengukur dalam satu arah dan satu wajah. (Azrilah Abdul Aziz et al. 2013). Instrumen yang mempunyai ciri kesamaran dan kekeliruan di kalangan responden perlu dilihat semula dan dimurnikan bagi memperolehi objektif instrumen untuk mengukur adalah kukuh dan boleh dicapai. Analisis Rasch menggunakan teknik Residual Principal Component Analysis (PCA) mampu mengesan kemampuan instrumen dalam satu dimensi yang seragam dengan tahap gangguan (noise) item yang boleh diterima. Linacre (2002) menggariskan bahawa nilai varians yang sebaik-baiknya adalah > 60%. Walau bagaimanapun, setiap konstruk yang ditunjukkan dalam varians kasar telah mencapai keperluan keseragaman instrumen iaitu sekurang-kurangnya 20%. Berdasarkan Jadual 8, sebanyak

**Detecting standardized residual correlation values**

To assess whether any items are confusing or overlap, the standard residual correlation (SRC) test should be used to guarantee that the instrument is devoid of confusion and missed objectives. If there are two items that record a value above 0.7, then it indicates a high correlation value because the characteristics are similar to each other and incorporate several other dimensions that are shared together. Thus, only one item is required for measurement. Based on Table 7, all items have a correlation value exceeding the value of 0.7. This means that respondents see the items found in this teacher competency construct as different things.

Table 7
*Largest Standardized Residual Correlation on Items*

LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT ITEM

| CORREL-ATION | ENTRY NUMBER | ITE | ENTRY NUMBER | ITE |
|---|---|---|---|---|
| .62 | 7 | D7 | 15 | D15 |
| .59 | 9 | D9 | 12 | D12 |
| .58 | 2 | D2 | 3 | D3 |
| .49 | 1 | D1 | 2 | D2 |
| .46 | 12 | D12 | 23 | D23 |
| .46 | 8 | D8 | 13 | D13 |
| .45 | 19 | D19 | 23 | D23 |
| .44 | 2 | D2 | 5 | D5 |
| .43 | 1 | D1 | 3 | D3 |
| .42 | 22 | D22 | 25 | D25 |
| -.53 | 1 | D1 | 15 | D15 |
| -.52 | 11 | D11 | 18 | D18 |
| -.52 | 4 | D4 | 23 | D23 |
| -.45 | 2 | D2 | 15 | D15 |
| -.44 | 16 | D16 | 23 | D23 |
| -.43 | 6 | D6 | 12 | D12 |
| -.43 | 2 | D2 | 22 | D22 |
| -.43 | 1 | D1 | 19 | D19 |
| -.43 | 6 | D6 | 15 | D15 |
| -.43 | 1 | D1 | 9 | D9 |

**Measuring Dimensional Uniformity**

Dimensional uniformity is crucial for assessing if an instrument can measure in a single direction and on a single face (Azrilah et al., 2013). Instruments that have characteristics of ambiguity and confusion among respondents need to be reviewed and refined to obtain the objective of the instrument to measure that is robust and achievable. Rasch analysis using the Residual PCA technique is able to detect the capability of the instrument in one dimension that is uniform with the level of interference (noise) of acceptable items.

According to Linacre (2002), the optimal value of variance is > 60%. The test consisted of 39 items that had a variance measured at 63.2% and showed that all items had reached a good value level of above 60.0% as set by the Rasch Model. While the variation value is unaccounted for, 1 is as high as 8.5 percent, which is well within the acceptable range of 15%.

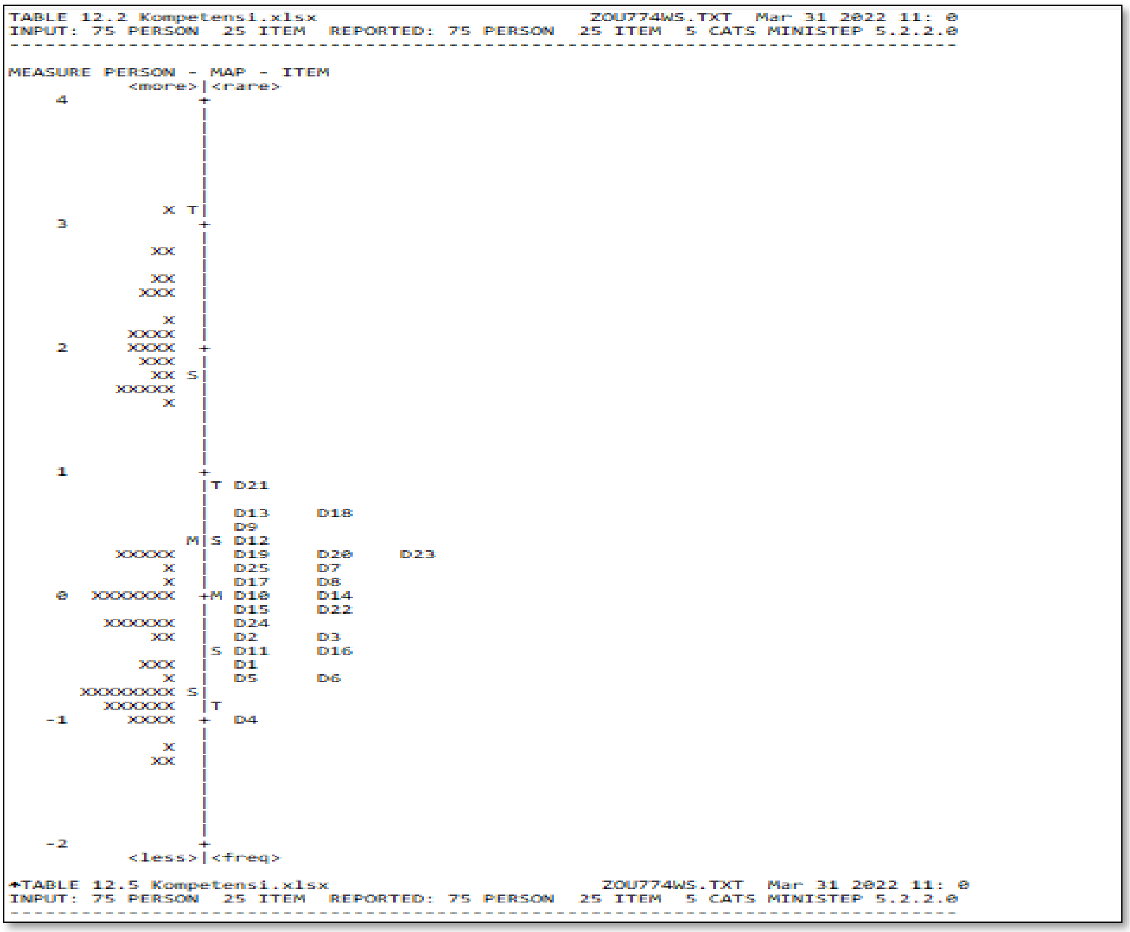
Table 8
*Unidimension: Standardized Residual Variance*

```
TABLE 23.0 Kompetensi.xlsx                       ZOU774WS.TXT  Mar 31 2022 11: 0
INPUT: 75 PERSON  25 ITEM  REPORTED: 75 PERSON  25 ITEM  5 CATS MINISTEP 5.2.2.0
--------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units
                                           Eigenvalue   Observed   Expected
Total raw variance in observations     =      68.0031 100.0%         100.0%
  Raw variance explained by measures   =      43.0031  63.2%          61.7%
    Raw variance explained by persons  =      22.2649  32.7%          31.9%
    Raw Variance explained by items    =      20.7381  30.5%          29.8%
  Raw unexplained variance (total)     =      25.0000  36.8% 100.0%   38.3%
    Unexplned variance in 1st contrast =       5.7708   8.5%  23.1%
    Unexplned variance in 2nd contrast =       2.6233   3.9%  10.5%
    Unexplned variance in 3rd contrast =       2.3943   3.5%   9.6%
    Unexplned variance in 4th contrast =       2.0774   3.1%   8.3%
    Unexplned variance in 5th contrast =       1.6536   2.4%   6.6%
```

**Item and Respondent Mapping**

Rasch analysis can provide a map of the item distribution that correlates to the respondents' abilities or tendencies. According to Bond and Fox (2007), this mapping demonstrates the relationship between pupils' talents and item difficulty levels. Respondents with high ability and high level of difficulty items will be located at the top of the scale, while respondents with low ability and items at a low level of difficulty will be located at the bottom. This is because measurements that use logit scales are shown from the easiest level to the most difficult level. The findings of the study, as shown in Figure 1, found Item D21, was the most difficult item to agree on with a measure of +1.35 logit. While Item D4 is the most easily agreed upon item with a measure of -2.15 logit, Values categorized as good and adequate are between +3.00 and -3.00 logit (Azrilah et al., 2013). The results meet the stated range, which is +2.29 to -1.22 logit.



```
TABLE 12.2 Kompetensi.xlsx                                  ZOU774WS.TXT   Mar 31 2022 11: 0
INPUT: 75 PERSON  25 ITEM  REPORTED: 75 PERSON  25 ITEM  5 CATS MINISTEP 5.2.2.0
--------------------------------------------------------------------------------

MEASURE PERSON - MAP - ITEM
          <more>|<rare>
    4           +
                |
                |
                |
                |
                |
                |
                |
            X  T|
    3           +
           XX   |
                |
           XX   |
          XXX   |
                |
            X   |
         XXXX   |
    2    XXXX   +
          XXX   |
           XX  S|
        XXXXX   |
            X   |
                |
                |
                |
                |
    1           +
                |T D21
                |
                |  D13     D18
                |  D9
               M|S D12
        XXXXX   |  D19     D20     D23
            X   |  D25     D7
            X   |  D17     D8
    0 XXXXXXX   +M D10     D14
                |  D15     D22
       XXXXXX   |  D24
           XX   |  D2      D3
                |S D11     D16
          XXX   |  D1
            X   |  D5      D6
     XXXXXXXX  S|
       XXXXXX   |T
   -1    XXXX   +  D4
                |
            X   |
           XX   |
                |
                |
                |
                |
                |
   -2           +
          <less>|<freq>
+TABLE 12.5 Kompetensi.xlsx                                 ZOU774WS.TXT   Mar 31 2022 11: 0
INPUT: 75 PERSON  25 ITEM  REPORTED: 75 PERSON  25 ITEM  5 CATS MINISTEP 5.2.2.0
--------------------------------------------------------------------------------
```

Figure 1. Item and Respondent Mapping

**Measuring the validity of the best rating scale**

Scale calibration is an important element in any measurement system and data validity. A good scale will work and will form a category based on the response, which should increase in parallel with the increase in the directed scale. If the mean measure of the category is not significantly different and does not show any improvement, then the adjacent category (scale) should be included (if the difference is <1.4). If the category mean measures are not significantly different and do not show any improvement, then adjacent categories (scales) should be grouped (if the difference < 1.4). As for if the scale value (threshold) > 5, then the scale needs to be separated. In this study, a 5-point likert scale was used which is (1) very infrequent; (2) infrequent; (3) quite often; (4) often, and (5) very often. Further descriptions of each frequency scale are also included. This type of ranking scale provides an opportunity for respondents to mark levels based on their perceptions (Najib, 2003).

The findings of the study from Table 9 indicate that the absence of differences exceeds or is less than the value of 1.4 on the part of the Andrich Threshold. Thus, the scales expected and used for each construct are appropriate, and the scales do not need to be separated or summarized. The Observed Average section also proves that the response pattern is also considered normal because it is found that there is a regular increase from negative to positive values for all three constructs. The validity of this scale can show and even show that the scales that were chosen for each construct are the right ones and that the chances of getting a response are spread out evenly (equally) between the scales.

Table 9
*Rating Scale Calibration Structure for Teacher Competency Constructs*

SUMMARY OF CATEGORY STRUCTURE. Model="R"

| CATEGORY LABEL | SCORE | OBSERVED COUNT | % | OBSVD AVRGE | SAMPLE EXPECT | INFIT MNSQ | OUTFIT MNSQ | ANDRICH THRESHOLD | CATEGORY MEASURE | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 362 | 19 | -1.02 | -.91 | .76 | .78 | NONE | ( -1.93) | 1 |
| 2 | 2 | 223 | 12 | -.46 | -.61 | 1.36 | 1.15 | -.28 | -.86 | 2 |
| 3 | 3 | 225 | 12 | -.17 | -.16 | .90 | .95 | -.42 | -.17 | 3 |
| 4 | 4 | 509 | 27 | .90 | .76 | .91 | 1.02 | -.57 | .70 | 4 |
| 5 | 5 | 556 | 30 | 1.78 | 1.89 | 1.20 | 1.09 | 1.27 | ( 2.48) | 5 |

**Conclusion**

Therefore, based on the pilot study conducted, it can be concluded that validity and reliability are among the important aspects that need to be conducted before developing a new instrument. This is because an authentic instrument will be established to enable measurements to be measured with the expected accuracy. To be sure, the refined instruments were found to be able to show better reliability characteristics.

The results of the evaluation of this instrument using the Rasch measurement model explain that this TECOMS instrument has a high level of quality in terms of validity and reliability and can be used in this study. The implications of the evaluation and examination results on this instrument allow the researcher to analyze the level of competence of teachers in small schools. This is the researcher's first step in helping others in producing teacher competency instruments.

**Acknowledgments**

We thank all our anonymous informants in this study and the reviewers for many insightful suggestions. Thanks are also given to the Faculty of Education, Universiti Kebangsaan Malaysia (UKM), and the Malaysian Ministry of Education (MOE) for their funding and scholarship in conducting this study.

**Corresponding Author**
Aida Hanim A. Hamid
Center for the Study of Leadership & Educational Policy, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia
Email: aidahanim@ukm.edu.my

**References**

Azrilah, A. Z., Saidfudin, M., & Azami, Z. (2013). *Fundamentals of the Rasch Measurement Model: Scale Formation & Measurement Structure*. Bangi: Universiti Kebangsaan Malaysia Publishers.

Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd). Mahwah, NJ: LEA.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences*. Routledge.

Carroll, A., York, A., Fynes-Clinton, S., Sanders-O'Connor, E., Flynn, L., Bower, J. M., Forrest, K. & Ziaei, M. (2021). The Downstream Effects of Teacher Well-Being Programs: Improvements in Teachers' Stress, Cognition and Well-Being Benefit Their Students. *Front Psychol*. 1;12:689628.

Davis, L. L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research. 5*, 194-197.

Emil, S., & Sedat, G. (2022). Socioeconomic achievement gaps and the role of school leadership: Addressing within- and between-school inequality in student achievement. *International Journal of Educational Research*, 112.

Haslinda, H., Noriam, S. Y., & Radzani, M. A. R. (2017). Teacher competence and its relationship to student co-curricular achievement in secondary schools. *Education Symposium: The Perspective of Risalah An-Nur* (SPRiN2017).

Linacre, J. M. (2002). What do Infit and Outfit, mean-square and standardized mean?. *Rasch Measurement Transactions*, 16, 878.

Linacre, J. M. (2007). *A user's guide to WINDTEPS Rasch-model computer programs*. Chicago, Illinois: MESA Press.

Losius G., Khairuddin, M. A., & Rosy, T. (2018). Teacher pedagogical competence and academic achievement in low-achieving schools. *Curriculum Journal, Curriculum Development Division, Ministry of Education Malaysia*, Volume 3, 80-91.

Mazdi, M., Jabil, M., & Rosmiza, M. Z. (2014). Peeling dropouts of Orang Asli Malaysian students: A survey into school accessibility issues. Geography: *Malaysian Journal of Society and Space*. 10 (2), 189-198.

Ministry of Education. (2009). Malaysian Teacher Standards. Putrajaya: Teacher Education Division.

Ministry of Education. (2012). *Malaysia Education Blueprint 2013-2025*. Putrajaya.

Ministry of Education. (2012*). School performance improvement tools for underserved schools*. Putrajaya.

Ministry of Education. (2017). *MOE professional circular No. 9/2017: Guidelines for the implementation of blended classes with multi -grade teaching and learning for the implementation of blended classes*. Putrajaya.

Ministry of Education. *Executive summary of underserved school roadmap*. Putrajaya.

Ministry of Education. (2019). *Minutes of human resource development panel* (PPSM) meeting No. 1 of 2019. Putrajaya.

Ministry of Education. (2020). *Malaysian Teacher Standard 2.0*. Putrajaya: Teacher Professionalism Division.

Ministry of Education. (2021). https://www.moe.gov.my/statistics-menus/statistics-number-school-teachers Downloaded on 15 October 2021.

Kashfi, M. J. (2011). *Quick & winstep measurement introduction manual. Measurement and evaluation in education*. Faculty of Education of the National University of Malaysia. Bangi.

Najib, M. A. G. (2003). *Design of an Educational Survey Questionnaire*. Universiti Teknologi Malaysia Skudai.

Nitce I. M. M. (2020). The influence of principals 'instructional leadership practices on teachers' self-efficacy in Underserved Schools. Doctor of Philosophy Thesis. Universiti Kebangsaan Malaysia.

Normiati & Abdul Said. (2019). The influence of teacher teaching on the quality of teacher teaching. Malaysian *Journal of Social Sciences and Humanities* (MJSSH), 4 (2), pp. 30-42. doi: 10.47405/mjssh.v4i2.189.

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 and Health* 30: 459-467.

Sharma, R., Sharma, M. S., & Agarwal, P. K. (2021). Effect Of Stress At Workplace And Its Management. *The International Journal of Business and Management Research*. 11(1), 1-12.