Jurma : Jurnal Program Mahasiswa Kreatif
Universitas Ibn Khaldun Bogor, Indonesia
Vol. 6, No. 2, Desember 2022, pp. 267~275
ISSN: 2615-8019, DOI: 10.32832/pkm
http://pkm.uika-bogor.ac.id/index.php/pkm-p

Psychometric Analysis of an Instrument Evaluating Online Efl Teaching Presence

Arum Kusumawardani [1], Herri Mulyono [2]
Universitas Muhammadiyah Prof. DR. HAMKA

Article Info

Article history:

Received June 16, 2022
Revised August 25, 2022
Accepted November 9, 2022

Keywords:

Online Teaching Presence
Pre-Service Teachers
Psychometric
Rasch Analysis

ABSTRACT

This research intended to examine the psychometric characteristics of the Indonesian version of the instrument used in evaluating the performance of pre-service teachers when practicing online teaching which was adapted from Howard. This quantitative study used convenience sampling technique by conducting a survey to collect respondent data. A total of 13 questionnaire items were developed using a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). Data were collected from 88 students who were carrying out online teaching practices or who had finished carrying out these activities. Several statistical analyzes were performed using Rasch analysis to assess unidimensionality, reliability, item statistic, person and item measure, and wright maps. The findings of this study revealed that the items in Howard's Indonesian version of the questionnaire are less valid and less reliable with the Indonesian sample, and the items do not retain the psychometric characteristics of the original scale because it failed to meet the minimum criteria of 0.80 logits. Implications and recommendations were discussed and offered with reference to the findings.

Corresponding Author:

Herri Mulyono
Universitas Muhammadiyah Prof. DR. HAMKA
Email: hmulyono@uhamka.ac.id

INTRODUCTION

Teaching practicum has been defined by several researchers. Richards & Crookes (1988) and Kim (2020) have defined teaching practicum similarly. According to Richards & Crookes (Richards & Crookes, 1988) teaching practicum is an opportunity to gain experience in order to improve the knowledge and teaching skills of education students who are studying to become teachers. Furthemore, Kim (Kim, 2020) teaching practicum provides hands-on experience for teacher education students to teach students in the classroom. Experience of teaching students directly in class is very much neede for pre-service teachers to add insight related to teaching activities as well as train and improve their teaching skill in real condition. For example, skill in communication, collaboration, critical thinking, managing skill, networking, leadership, and problem solving. A study by El Kadri & Roth (2015) teaching practice is often seen as a necessary activity to civilize teacher education students by offering opportunities to experience teaching experiences before joining the teacher community. Teacher education students are given the opportunity to explore experiences as teachers by conducting practical activities before they join as teachers officially. For instance, almost all universities that have education faculties require their students to carry out teaching practice activities in order to gain experience and improve their teaching skills even though it is done online. It is supported by Richards & Crookes (1988), which says that teaching practicum is a required course for all students in many teacher education programs since it is an integral aspect of teacher professional development. Thus, teaching practicum

can be defined as a program for teacher education students to gain experience, increase knowledge, and train teaching skills that must be taken before becoming professional teachers.

This study attemps to examine the Online Teaching Presence instrument which was adapted from Howard. In particular, the current study will assess the validity and consistency of the instrument using Rasch modeling. Rasch analysis was performed using WINSTEP (version 4.4.1) to evaluate quantitative data. According to Yu, (2020), as cited in Mulyono et al., (Mulyono et al., 2020), the psychometric technique that is widely used by researchers or instrument developers in monitoring instruments by assessing the ability of people and the difficulty of items simultaneously is Rasch analysis. This study includes several assessments of the reliability of the separation of items and persons; rating scale; wright map; and item bias. In the literature, the use of the current Rasch analysis to validate the Onile Teaching Presence Scale adapted from Howard has not been found. Therefore, providing empirical evidence referring to psychometric scales using Rasch analysis is very important in providing the current literature. More importantly this study will provide a new investigation that is rarely explored in the literature on the pre-service teachers performance in online teaching.

REVIEW OF LITERATURE

Teaching practice is not only teaching face-to-face, but also in an online context. Especially during these Covid-19 pandemic, the previous conventional teaching changed to online teaching. Online teaching is teaching practice that is entirely done online (Baran, Correia, & Thompson, 2011) and a new way to interact with students in the teaching and learning process (Coman, Țîru, Meseșan-Schmitz , Stanciu, & Bularca, 2020). Teaching practice usually begins with observing the teaching teacher, then gradually takes over responsibility for teaching students in the class under the supervision of a cooperating teacher (Richards & Crookes, 1988). However, since many schools are unable to provide teaching practicum during COVID-19 pandemic, pre-service teachers must undertake "internal" internships, with other teacher candidates acting as their students (Brinia & Psoni, 2021) or by doing online teaching practice.

Several studies have found about online teaching practicum. The findings by Carrillo & Flores (2020) highlight the necessity for an approach to online education pedagogy that incorporates technology to aid teaching and learning. In addition, a study conducted by Rhode et al., (2017) highlighted a broad disagreement among teacher educators on whether online classes are an effective technique for preparing pre-service teachers, as well as some of the complexities of online teaching educators in terms of teacher identity, role modeling, and pedagogy enforcement. Another study conducted by Baran et al., (Baran et al., 2011) showed that pre-service teachers will be able to participate in community practice and transform their teaching by expanding on their knowledge and social practices by introducing collaborative working groups, community building, and group discussions into professional development programs and preserving continuity. Therefore, online teaching preparation, guidance, and support for pre-service teachers are indispensable to know how to build their personas through online pedagogy.

Initially the Online Teaching Presence (OTP) instrument was made by Arbaugh et al., (2008) which consisted of 34 items. Then the instrument was developed by Gurley (Gurley, 2018) into 13 items where Online Teaching Presence (OTP) for clear instruction consisted of 4 items (Q1 to Q4); student feedback and assessment consists of 4 items (Q5 to Q8); and cognitive activation consisted of 5 items (Q9 to Q13). From the 13 items, it shows that the Crombach Alpha is at the level of 0.852, which means it is at a high level. This instrument was later adapted by Howard et al., (Howard et al., 2021) to assess the readiness of teachers in the transition from conventional to online teaching. First, Howard et al (Howard et al., 2021) used Confirmatory Factor analysis (CFA) to test the psychometric quality of the scale. It then uses Latent Profile analysis (LPA) to offer a model-driven and more flexible approach to group identification than cluster analysis and represents a human-centric approach that uses a set of profile indicators to identify similar groups within a sample. The current study is carried out by adapting an instrument that has been developed by Howard et al., (Howard et al., 2021) using the Rasch model which has never been used to test the Online Teaching Performance instrument.

METHOD

This research using the instrument adapted from Howard et al., (2021). There are five demographic information scales and one construct used in the instrument; namely Online Teaching Presence (OTP). The OTP was used to examine teachers' impressions of their online presence, and it contains of 13 items were labeled Q1 to Q13. It through three important aspects: OTP for clear instruction consists of 4 items (Q1 to Q4); student feedback and assessment consist of 4 items (Q5 to Q8); and cognitive activation consists of 5 items (Q9 to Q13). Furthermore, a total of 13 questionnaire items will be developed using a 5-point Likert scale: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree.

ISSN: 2615-8019

This study on the performance of pre-service EFL teachers in online teaching uses a convenience sampling technique. The survey was conducted to collect respondent data. The participants of this research consisted of 88 students who are currently studying at several universities in Indonesia. Respondents were students who were doing internships at schools as much as 27% (n=24) or 73% who had finished doing internships at schools (n=64). From the total of 88 respondents, there were 76 female respondents (86%) and 12 male respondents (14%). Pre-service teachers range in age from less than 22 to more than 25 years. Pre-service teachers with varied levels of school taught and technology abilities were also among those who took part.

The WINSTEP application (version 4.4.1) was used in the Rasch analysis to evaluate quantitative data including reliability, RASCH model fit, variance, item bias, and item-person mapping (Wahid et al., 2021). In full, this research includes several assessments of an item, and person fit through the Outfit statistic and through the mean square before Z-standardized which will be carried out to overcome outliers; dimensionality; reliability of separation of items and person; scoring scale; mapping of items and person; and item bias. First, a questionnaire in the form of a Google form link is distributed via social media, such as Whatsapp, Instagram, and Twitter. Prior to completing the questionnaire, participants were asked to fill in demographic information (age, gender, school level, technology skills, and internship status). Then, as many as 263 data which are the answers of the respondents were converted into excel form. After that, the incoming data filtered by retrieving only the data from respondents who were currently and have had online internships at schools. Currently, there are a total of 116 teachers who are conducting and have completed online internships at school.

The range of acceptable apparel statistic values for MNSQ is 0.5 to 1.5 and for ZSTD between -2.00 and 2.00. Dimensional assessment is perfomed by evaluating Rasch Principle Component Analysis (PCA), and bias items are carried out by assessing Item Differential function (DIF) with criteria: (DIF contrast > 0.5) and (probability < 0.05). Statistical analysis was performed on the data set to address potential outliers that might violate data analysis. Of the 116 samples, 28 were found to be outliers, and the remaining 88 samples are known to be able to be analyzed in the second round using the WINSTEP application.

According to Bond (2015), Rasch analysis can be used to calibrate the ability person parameter and the difficulty item parameter on the same unidimensional scale. Parameters are expressed in logits which is a natural algorithm of odd ratios. In Rasch's measurement, when one's ability was greater than the difficulty of the item, the bigger the difference, the higher the probability of answering the item correctly. When one's ability is lower than the item's difficulty, the bigger the difference, the lower the probability of answering the item correctly.

Rasch analysis using WINSTEP (version 4.4.1) to evaluate the quantitative data. The study includes several assessments of an item, and person fit through Outfit statistics and through mean square before Z-standardised that will be performed to address outliers; dimensionality; item and person separation reliability; rating scale; item and person mapping; and item bias. The acceptable values of the outfit statistics range between 0.5 and 1.5 for MNSQ; and for ZSTD between -2.00 and 2.00. Dimensionality assessment is perfomed by assessing the Rasch Principle Component Analysis (PCA), and item bias was done by assessing Differential Item functioning (DIF) with the criteria > 0.5 for DIF contrast and < 0.05 for probability.

FINDINGS AND DISCUSSION

Findings

The results of the Rasch analysis provided information about the characteristics of the test as an instrument and information on the performance of EFL pre-service teachers in the implementation of online teaching. On the characteristics of the instrument, the first test criterion were the degree of fit of the data model, namely the extent to which the assessment data is in accordance with the requirements of the Rasch model. Rasch analysis provided fit statistics to evaluate the data empirically. In addition, other statistics such as item and person reliability, and item and person separation index were examined and reported to inform the quality of the test.

Adapting the table from the previous research by Ling Lee (2020), the results of the Rasch analysis were reported in the summary below.

Table 1. Summary of Rasch measurement model on Online Teaching Presence

Parameter (with quality criteria)	Online Teaching Presence
Model fit: Summary of items	
Item mean in logits (criteria: 0.0 logits)	.00, SD = .47
Item reliability	.75
Item separation reliability (criteria: good, 0.81-0.90; very good, 0.91-0.94; excellent, >0.94)	.75
Item model fit MNSQ range extremes (criteria: good, 0.5-1.5; very good, 0.71-1.4; excellent, 0.77-1.3)	Infit .68 – 2.12 Outfit .56 – 2.16
Item separation index (criteria > 3)	1.74
Separate item strata = [(4 x separation index) + 1]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	2.65 ≈ 3 levels
Model fit: Summary of persons	
Person mean in logits (criteria: 0.0 logits)	3.89, SD = 3.89
Person reliability	.96
Person separation reliability (criteria: good, 0.81-0.90; very good, 0.91-0.94; excellent, >0.94)	.96
Person separation index (criteria > 2)	4.78
Separate Person strata = [(4 x separation index) + 1]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	6.70 ≈ 7 levels
Dimensionality	
Raw variance in data explained by measure (criteria: > 20%)	68.9%
PCA eigenvalue for first contrast (criteria: > 2.0 indicates presence of another dimension; ≤ 2 supports unidimensional scale)	3.1
Unexplained variance in 1st-5th contrast of PCA of residuals (criteria: good, 5-10%; very good, 3-5%; excellent, <3%)	2.4% - 7.5%

Note: Table is adapted from Ling Lee et al (2020)

Unidimentionality of the items

Unidimensionality was done by assessing the Rasch Principal Component Analysis (PCA) on a global scale and all subscales to find out that the items in the instrument only measure one construct. In other words, when the Rasch dimension was extracted, this looked to see whether there was more than one dimension that might explain the data.

The PCA eigenvalue criteria for the first contrast was more than 2 logit, while the criteria for raw variance in data explained by measure is more than 20%. In other words, it can be said that it was unidimentionality if the eigenvalues of PCA and raw varience showed results that did not match the criteria, namely < 2 logit and < 20%.

ISSN: 2615-8019

In this analysis, the PCA showed the eigenvalue of the first contrast above the criteria to be 3.1 logit, whereas 68% of the variance in the data was described by the Rasch measure which also did not meet the unidimensional criteria. This indicated that the Online Teaching Presence instrument shows mismatch with the Rasch model, and the construct was not measured in a unidimension. Thus, it can be seen that OTP indicated the existence of other dimensions.

Person and Item Reliabilities

The result showed that the item separation reliability for Online Teaching Presence scale was .75 logit (fair). It was also known that the person separation reliability of the OTP scale at .96 logits with "excellent" criteria. It was clear from this discovery that the reliability of the items quality on the instrument was "fair", and the respondents' answers was "excellent". Furthermore, it was known that the item separation index of OTP was 1.30 logit. That was, at least one level of item difficulty can be classified of the items on the instrument. It can be said that all items in the questionnaire either difficult, moderate, or easy. In addition, Item separation index for OTP was 1.74 logits. OTP scale were unfit the criteria because the logit for OTP was under criteria (> 3).

Item Statistics

For each of the 13 items, item statistics such as item difficulty, standard error of measurement, point-measure correlations, infit and outfit statistic were provided in the table below. Except for item number 10 (infit - outfit MnSq at 2.12 and 2,16), all of the 12 items had a fit statistic ranging from 0.5 to 1.5 Khine (2020).

Table 2. Item measure, standard error, fit, and point-measure correlation

Item	Difficulty	Standard Error	Infit		Outfit		PTME corr.
			MnSq	ZStd	MnSq	ZStd	
1	.29	.25	1.03	.21	.89	-.49	.85
2	-.03	.26	1.17	1.00	1.04	.23	.86
3	-.10	.26	.87	-.69	.74	-1.27	.89
4	-.72	.27	1.07	.40	.94	-.16	.89
5	.10	.26	.80	-1.16	.77	1.14	.89
6	-.72	.27	.72	-1.60	.73	-1.21	.91
7	-.58	.26	1.06	.40	1.09	.46	.88
8	-.03	.26	.92	-.39	.82	-.83	.89
9	-.10	.26	.79	-1.23	.73	-1.29	.90
10	1.41	.24	2.12	6.23	2.16	5.63	.72
11	-.10	.26	.68	-2.00	.56	-2.40	.91
12	.03	.26	.79	-1.25	.86	-.65	.89
13	.54	.25	.77	-1.54	.71	-1.64	.90
Mean	.00	.26	.98	-.1	.93	-.4	

SD	.54	.01	.36	2.0	.38	1.9

Person and Item Measures

In the table above, it demonstrated that the item difficulty range from -.72 logits to 1.41 logits. The standard error of measurement mostly showed at .26 logits and the other item such as Q10 = .24; Q1 and Q13 = .25; Q4 and Q5 = .27. According to Khine (M. S. Khine, 2020) the most difficult item (>2.00 logits) was whereas the easiest item (<−2.00 logits). In this table, it can be seen in the difficulty column that the highest logit value wa at 1.41 where it was the most difficult item but has not yet reached the estimation criteria. It means that Q10 = 1.41 logits was not yet classified as the most difficult item or it can be said that the item was classified as a moderate item. Meanwhile, the lowest logit value was at -.72 where it was the easiest item but has not yet reached the estimation criteria. That was, Q6 = -.72 logits was not yet classified as the easiest item or it can be said that the item was classified as an average item. In other words, it can guarantee that all items belong to the moderate questionnaire item group.

Item-Person Map

Wright map was used to determine the distribution of person levels and difficulty of items on the same scale. The distribution of the measured level of person was shown on the left side of the map in the picture below, while the distribution of the difficulty level of item was shown on the right side of the map. At the highest level it was known for the items that were the most difficult to agree on, while at the lowest level were the items that were the easiest to agree with by the respondents. The result showed that the easiest item was at -.72 logits and the hardest item was around 1.41 logits.

```
MEASURE   Person - MAP - Item
              <more>|<rare>
   11            X  +
                   T|
                    |
   10         XXXX  +
                    |
                X   |
    9               +
             XXX    |
              XX    |
    8         XX    +
               X    |
               X S  |
    7              +
               X    |
            XXXXX   |
    6           X   +
                    |
            XXXXX   |
    5 XXXXXXXXXXX   +
                    |
         XXXXXXXXX  |
    4         XXXX M+
                    |
          XXXXXX    |
    3       XXXXXX  +
              XXX   |
               XX   |
    2          XXX  +
```

ISSN: 2615-8019

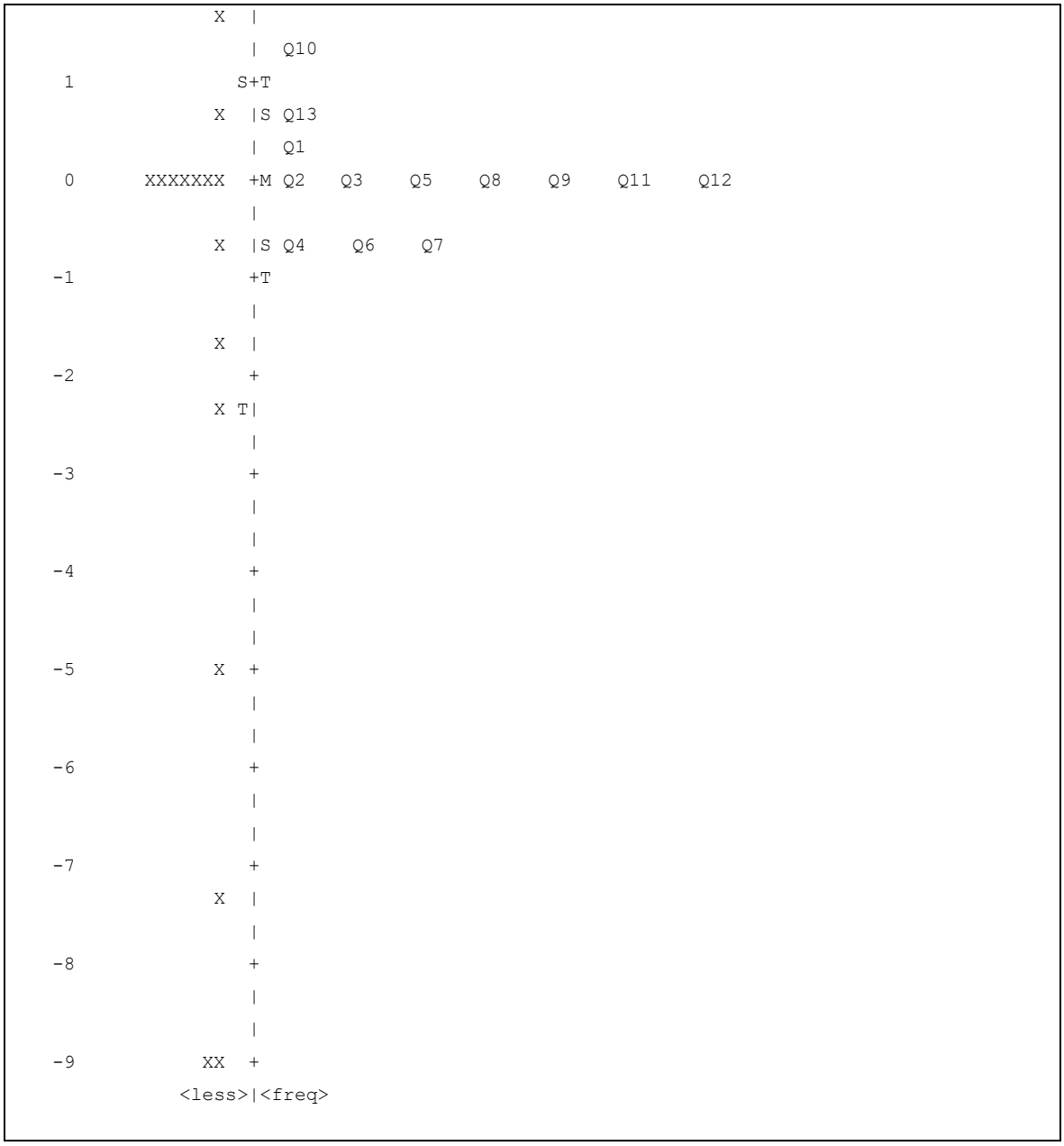

Figure 1. Wright Maps

It can be seen that the item Q10 = 1.41 logit of Online Teaching Presence was shown as the most difficult item or the least selected item for approval by the respondents. It was represented as the most difficult item to agree on or the least chosen by only a few responders. For instance, item Q10 (Activities in my online learning strengthen togetherness among students) was the most difficult item for responders to agree on. According to the statement, the majority of respondents believe that activities in their online learning do not improve student collaboration, with 73 responding in the negative.

While the questions Q7 (-.58 logit), Q6 (-.72 logit), and Q4 (-.72 logit) proved to be the easiest or it can be said that many respondents agreed. For example, respondents rated item Q4 “I clearly communicate schedules and deadlines related to online learning activities” as the most agreeable. This means that the majority of pre-service teachers clearly communicate schedules and deadlines related to online learning activities. Item Q6 “I help students understand online learning topics and materials”. This showed that most pre-service teachers assist students in understanding online learning topics and materials. Item Q7 “I help students stay engaged and participate in online learning”. It can be seen that most pre-service teachers help students to stay involved and participate in online learning.

Item Fit

Except for Q10, which has Infit MnSq = 2.12 and Outfit MnSq = 2.16, all items had infit and outfit statistics less than 2.0. This showed that there was no misfit in all of the items, except the item Q10. The Rasch analysis also reported information on other statistical indices, such as the point-measure correlation, which was a measure of the relationship between responses to a single item and the total test score, and positive scores were thought to support the internal coherence of items in contributing to test scores. All items showed a correlation value from 0.72 to 0.91. These results indicated that the items in the instrument were internally coherent and can be understood by respondents.

Discussion

This paper reports on an analysis of preservice teacher performance data using the Rasch model. This research report provides a description of the performance of pre-service teachers, especially for those who are currently carrying out online teaching practices or who have finished carrying out these activities. Information on the results of this instrument assessment can be used for pre-service teachers in improving their performance when practicing online teaching. The results provide a comparison between the item difficulty level and the pre-service teacher performance level. Thus, pre-service teachers can find out what performance they find difficult to implement and can also find out at what level of performance they are.

Pre-service teachers found some items that they found difficult to answer, as previously reported, they found it more difficult to agree with the statement "My online learning activities strengthen the togetherness of students". This means that they feel that activities in online learning do not strengthen togetherness between students. This is due to the absence of face-to-face interaction so that there are limitations in increasing togetherness. However, this can be overcome by making activities that can strengthen togetherness between students, such as holding group discussion forums or making joint projects in one group so that togetherness can increase because of the interaction when they work together.

Overall, the assessment of the instrument's performance of pre-service teachers when implementing online teaching practices can be useful for prospective teachers to improve their performance and make it an evaluation material for future online teaching practices. The suggestions given can also be used to add references in improving their performance.

CONCLUSION

Rasch modeling was employed in this paper to test 13 pre-service teacher performance items in online teaching. According to pre-service teachers' response to the item questionnaire, assessment information is provided by psychometric analysis that is useful for evaluating the quality of the instrument in measuring the performance of pre-service teachers in online teaching. In summary, the reliability test showed that the items were less reliable because it failed to meet the minimum criteria of 0.80 logits.

REFERENCES

[1]. Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. Internet and Higher Education, 11(3–4), 133–136. https://doi.org/10.1016/j.iheduc.2008.06.003

[2]. Baran, E., Correia, A.-P., & Thompson, A. (2011). Transforming online teaching practice: critical analysis of the literature on the roles and competencies of online teachers. Distance Education, 32(3), 421–439. https://doi.org/10.1080/01587919.2011.610293

[3]. Bond, T. G. (2015). Applying The Rasch Model : Fundamental Measurement in the Human Sciences. In Taylor & Francis.

[4]. Carrillo, C., & Flores, M. A. (2020). COVID-19 and teacher education: a literature review of online teaching and learning practices. European Journal of Teacher Education, 43(4), 466–487. https://doi.org/10.1080/02619768.2020.1821184

[5]. El Kadri, M. S., & Roth, W. M. (2015). The teaching practicum as a locus of multi-leveled, school-based transformati. Teaching Education, 26(1), 17–37. https://doi.org/10.1080/10476210.2014.997700

[6]. Gurley, L. E. (2018). Educators' preparation to teach, perceived teaching presence, and perceived teaching presence behaviors in blended and online learning environments. Online Learning, 22(2), 197–220.

[7]. Howard, S. K., Tondeur, J., Siddiq, F., & Scherer, R. (2021). Ready, set, go! Profiling teachers' readiness for online teaching in secondary education. Technology, Pedagogy and Education, 30(1), 141–158. https://doi.org/10.1080/1475939X.2020.1839543

[8]. Khine, M. S. (2020). Rasch Measurement (M. S. Khine (ed.)). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-1800-3

[9]. Kim, J. (2020). Learning and Teaching Online During Covid-19: Experiences of Student Teachers in an Early Childhood Education Practicum. International Journal of Early Childhood, 52(2), 145–158. https://doi.org/10.1007/s13158-020-00272-6

[10]. Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A Rasch analysis. European Journal of Preventive Cardiology. https://doi.org/10.1177/2047487320902322

[11]. Mulyono, H., Saskia, R., Arrummaiza, V. S., & Suryoputro, G. (2020). Psychometric assessment of an instrument evaluating the effects of affective variables on students' WTC in face-to-face and digital environment. Cogent Psychology, 7(1), 1823617.

ISSN: 2615-8019

https://doi.org/https://doi.org/10.1080/23311908.2020.1823617

[12]. Rhode, J., Richter, S., & Miller, T. (2017). Designing Personalized Online Teaching Professional Development through Self-Assessment. TechTrends, 61(5), 444–451. https://doi.org/10.1007/s11528-017-0211-3

[13]. Richards, J. C., & Crookes, G. (1988). The Practicum in TESOL. TESOL Quarterly, 22(1), 9. https://doi.org/10.2307/3587059

[14]. Wahid, R., Halim, S., & Halim, T. (2021). Incorporating Creativity and Communication Skills among the Students of Media Department. TESOL International Journal, 16(6.1).

[15]. Yu, C. H. (2020). Objective measurement: How Rasch modeling can simplify and enhance your assessment. In M. Khine (Ed.), Rasch Measurement: Applications in Quantitative Educational Research (pp. 47–73). Springer. https://doi.org/10.1007/978-981-15-1800-3_4