



Cultivating distance learning: Developing and validating online teacher engagement scale using Rasch model

Ika Febrian Kristiana¹, Ermida Simanjuntak², Bambang Sumintono³, Unika Prihatsanti¹, Crossita Nabila Shofiefany¹

¹Faculty of Psychology, Universitas Diponegoro, Semarang, Indonesia

²Faculty of Psychology, Universitas Katolik Widya Mandala, Surabaya, Indonesia

³Faculty of Education, Universitas Islam Internasional Indonesia, Depok, Indonesia

ABSTRACT

Background: Teacher engagement plays a crucial role in determining the success of online learning, particularly as digital education becomes increasingly prevalent. However, instruments to accurately assess teachers' engagement in online learning contexts remain limited.

Purpose: This study aimed to develop and validate a measurement scale to assess teacher engagement in online learning.

Method: The study involved 411 teachers from elementary to middle schools across various regions in Indonesia. The initial version of the Online Teacher Engagement Scale (OTES) consisted of 40 items measured using a five-point Likert scale. Data were analyzed using the Rasch model.

Findings: The analysis demonstrated that the OTES possessed strong psychometric properties, with a person reliability index of .89 and an item reliability index of .99. The 32 items were valid after passed some item tests criteria in the Rasch analysis. The explained raw variance of 38% suggested that the OTES measured the intended construct adequately, although five items displayed gender-based DIF, leading to their exclusion.

Implication: OTES has broad potential as an evaluation instrument for schools, universities, and teacher training institutions in monitoring the level of teacher involvement during online learning.

KEYWORDS

Online learning;
psychometric study; Rasch
model; teacher engagement;
validation

ARTICLE HISTORY

Received 24 January 2026

Revised 14 August 2026

Accepted 10 September 2026

Introduction

In recent years, the learning process in schools has experienced rapid changes, along with developments in innovation and technology (Agasisti et al., 2023). To optimize the learning process in schools, this technology generally uses Information and Communication Technology (ICT) in the form of student learning instruments (Lin et al., 2023). An example of ICT usage in the school learning process is the learning management system (LMS), which is a server-based or web-based software program that consists of several databases containing information about users (students and instructors), courses, educational content, and learner outcomes that facilitate the learning process for students without being limited by time and place (Turnbull et al., 2020). The online learning situation, initially only in certain schools, later became common in most schools worldwide due to the COVID-19 pandemic (Sarosa, 2022). The successful implementation of online learning in schools depends on several factors, including students' and teachers' perspectives (Das, 2023). Teachers themselves play an essential role in the success of students' online learning (Leite et al., 2022; Sidi et al., 2023). A good teacher in an online learning situation has certain characteristics, including being able to design learning tasks that are appropriate for students (Archambault et al., 2022); becoming a facilitator and motivating student (Richardson et al., 2016); and understanding the technical matters required, such as how to use video conferencing and learning software, and operating learning technology tools (Corry & Stella, 2018; Grammens et al., 2022).

ICT integration offers numerous benefits to education, including fostering innovative teaching methodologies, enhancing student engagement, and promoting educational equity (Anastasopoulou et al., 2024). The adoption of digital tools enables educators to create dynamic

CONTACT Ika Febrian Kristiana ✉ ikafebriankristiana@lecturer.undip.ac.id 📠 Jl. Prof. Mr. Sunario, Tembalang, Semarang, Indonesia

Copyright © 2026 by The Authors. Published by Faculty of Psychology, Universitas Diponegoro. This an open-access article under the CC-BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

learning environments that accommodate a variety of learning preferences, increasing student engagement and motivation (Cueva & Inga, 2022). In exploring the specific context of rural environmental education, the integration of ICTs can be significantly effective when aligned with well-designed pedagogical strategies that focus on active pedagogy and critical digital competencies. These strategies can substantially improve sustainability learning and student engagement, highlighting the necessity for community collaboration and inclusivity (Baena-Navarro et al., 2024). However, the integration of ICT in education does not come without challenges. Issues such as the digital divide and insufficient ICT training for teachers present significant hurdles. Overcoming these requires focused efforts on capacity building, improving infrastructure, and tailoring educational interventions to the specific needs of communities (Ngodu et al., 2024). Preparing future educators involves developing their professional digital competence, as well as their ability to use ICT responsibly, ensuring that privacy, copyright, and ethical considerations are observed (Gudmundsdottir & Hatlevik, 2020). In summary, rethinking teacher engagement strategies in light of ICT advancements involves more than simply incorporating new technologies into the classroom. It requires a comprehensive approach that includes revising pedagogical frameworks, addressing infrastructural challenges, and ensuring that teachers are well-equipped to utilize ICT effectively and responsibly. By focusing on these elements, educators can enhance their engagement strategies and improve educational outcomes in the digital age (Anita, 2024).

Several studies have shown that teachers who are engaged in their duties demonstrate behaviors that support students' learning process in class (Bakker & Bal, 2010; Hakanen et al., 2019; Penttinen et al., 2020; Soininen et al., 2023). Examples of such engaged behaviors include exhibiting energy and enthusiasm while working and finding solutions to existing difficulties (Greenier et al., 2021). Teacher engagement is the concept of teachers' engagement in carrying out their tasks. This is related to the concept of work engagement proposed by Schaufeli et al. (2002), which refers to the engagement a person has in their work. Teachers who are engaged in their work direct all of their available abilities and resources to complete their duties (Bakker & Bal, 2010; Wal et al., 2014). Teacher engagement has frequently been measured using the Utrecht Work Engagement Scale (UWES) in previous studies (Bakker et al., 2014; Li et al., 2015). Schaufeli et al. (2002) developed the UWES as an instrument for measuring work engagement, comprising 17 items that span three dimensions: vigor, dedication, and absorption. Cronbach's α values on the UWES scale range between .80 and .90 for each dimension. The UWES Short Version consists of nine items with fairly good Cronbach's α values, between .85 and .92. Accordingly, this scale is known as the UWES-9 (Schaufeli et al., 2006). The UWES-9 instrument has been used to measure teacher engagement among Finnish teachers ($N = 2038$), and indicated a relationship between teacher burnout and teacher engagement (Hakanen et al., 2006).

In addition to the UWES-9, teacher engagement is also frequently measured using the Engaged Teachers Scale (ETS), an instrument developed by Klassen et al. (2013), which more specifically measures work engagement in the teaching profession. This scale comprises 16 items organized into four factors—cognitive engagement (CE), emotional engagement (EE), social engagement with students (SES), and social engagement with colleagues (SEC), with Cronbach's α values: CE = .85, EE = .89, SES = .84, and SEC = .85 (Klassen et al., 2013). ETS has been adapted and used to measure teacher engagement in several countries, including Australia (Perera et al., 2018), Turkey (Yerdelen et al., 2018), Brazil (da Silva Júnior et al., 2020), and Hong Kong (Ho et al., 2021). The results of these adaptation studies in several countries show that the ETS has internal consistency values for its 16 items in the medium to high range. In terms of construct validity, the fourth factor indicated a good model fit (fit: χ^2 , CFI, GFI, NFI, SRMR, RMSEA). Furthermore, findings from Perera et al. (2018) indicated a significant relationship between teacher engagement and job satisfaction. However, an engagement scale for teachers in an online learning context needs to be developed, especially in an era of widespread use of technology in education (Liang & Chen, 2012).

More than 90% of the measurements developed in psychology studies are based on the Classical Test Theory (CTT) approach. Nevertheless, this approach carries three fundamental theoretical limitations: (1) it presupposes a linear relationship between latent variables and

observed scores, a condition that seldom reflects the actual nature of behavioral constructs; (2) true scores cannot be directly estimated, or can only be approximated by relying on assumptions that are difficult to satisfy; and (3) parameters such as reliability, discrimination, location, or factor loadings are contingent upon the specific sample employed. To overcome this problem, Item Response Theory (IRT) offers a robust alternative for measuring continuous latent variables via categorical indicators (i.e., measurement variables), offering several benefits: it (1) does not require a linear relationship, (2) enables more accurate score estimation, (3) permits item parameters to be estimated independently of the sample, (4) allows researchers to choose items that best fit the intended model, and (5) extends and generalizes notions such as reliability and internal consistency (Chen et al., 2018; Fischer & Rose, 2016; Hambleton & Jones, 1993).

Based on IRT, Rasch analysis offers a particularly effective alternative for examining the psychometric properties of measurement instruments while accounting for response bias (Bradley et al., 2015). The original Rasch model was developed for use with items rated dichotomously (i.e., items marked as either true or false) and was based on Thurstone and Guttman's early work (Osterlind, 2009; Tennant & Küçükdeveci, 2023). In contrast to the CTT, which assumes that the standard error of measurement is uniform across test takers and dependent on the sample used, this approach assumes that measurement error differs from one individual to another and is independent of any particular respondent sample. Estimation of the latent traits under investigation relies on both person and item characteristics, with a person's ability and an item's difficulty placed on the same scale (logit), derived from response probabilities and nonlinear logarithmic functions. Accordingly, IRT-based analysis can be applied to assess whether item difficulty aligns with an individual's ability level on the latent trait in question. By matching item difficulty to a person's ability, the IRT allows the development of measures using more accurate score reliability with fewer test items. Further details regarding the theoretical foundations and computational procedures of Rasch analysis are provided by Embretson and Reise (2013), whereas a practical guide to its implementation using Winsteps software can be found in Bond and Fox (2012). The stages of developing the online teacher engagement scale (OTES) measurement tool follow the four-building-blocks approach based on the IRT (Wilson, 2023), which is as follows in Figure 1.

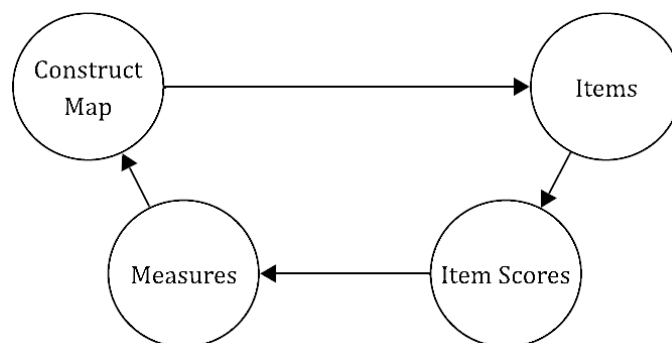


Figure 1. The Instrument Development Cycle Through the Four Building Blocks

Method

Participants

A total of 411 teachers originally participated in this study, consisting of 104 teachers in Phase 1 (pilot test) and 307 teachers in Phase 2. Following data-cleaning procedures in Phase 1, 9 non-fitting responses were removed, yielding a final analyzed sample of 402 teachers (95 in Phase 1 and 307 in Phase 2). The final sample comprised 88 males (21.9%) and 314 females (78.1%) from 18 elementary, middle, and high schools in Indonesia. The first language of the participants was Indonesian. All respondents provided consent to participate voluntarily after receiving an explanation of the research objectives on the scale form provided by the researcher. This research received ethical clearance from the Health Research Ethics Committee of the Faculty of Public Health, Universitas Diponegoro, under approval number 262/EA/KEPK-

FKM/2021. Data were collected online using Google Forms owing to its wider access and higher efficiency in distribution.

Data Collection and Analysis

The scale

At the construct map stage, the researcher determined the conceptual definition, measurement limits, and aspects of teacher engagement in online learning based on a literature review of the constructs of teacher engagement and the learning management system, and also synthesized the results of focus group discussions with teachers, students, and parents (see (Klassen et al., 2013; Grönlund et al., 2023; Wasson & Hansen, 2014). During the FGD phase, 20 teachers (with at least 6 months of online teaching experience, aged 30–45), 10 parents (with experience supporting their children's online learning), and 20 students (in grades 1–2 of junior high school, aged 11–13) were involved. The FGD data was analyzed thematically to gain an understanding of teachers' behaviors or efforts in online learning. Based on thematic analysis of the qualitative data, teacher engagement in online learning (Online Teacher Engagement) is defined as a teacher's work attitude, which is demonstrated by their dedication to devoting energy (vigor), thoughts (initiative), emotional closeness, collaboration, and adaptiveness to designing online learning, applying learning technology, and interacting with students.

Hence, there are five components of teacher engagement in online learning: vigor, initiative, emotional closeness, collaboration, and adaptiveness. These components emerged from a rigorous thematic analysis of FGDs within the construct mapping phase of scale development (Wilson's procedure). This methodological step ensures the new scale is grounded not only in theoretical traditions but also in lived experiences and contextual specifics, providing conceptual and empirical legitimacy. The process reaffirms that engagement, particularly in evolving contexts such as online learning, is multifaceted, dynamic, and relational. The measuring boundaries set in the online learning framework included pedagogical planning, application of learning technology, and student interaction. In the next step, the researcher created a blueprint and developed 40 items, where each component was represented by eight items, which were tested using a Likert-scale model with five response options (strongly disagree = 1, disagree = 2, somewhat disagree = 3, agree = 4, and strongly agree = 5). Scoring followed both favorable and unfavorable items (reverse scoring).

Table 1.

Instrument Reliability and Validity Analysis

The 1 st pilot study							
No	Component	Item reliability	Separation	INFIT MNSQ		OUTFIT MNSQ	
				Max	Min	Max	Min
1	Vigor	.98	7.63	2.24	0.2	2.34	0.2
2	Initiative	.97	5.51	1.61	0.57	1.63	0.67
3	Emotional Closeness	.99	8.82	1.57	0.5	1.60	0.46
4	Collaboration	.94	3.99	1.25	0.6	1.40	0.61
5	Adaptiveness	.98	7.36	1.63	0.48	1.59	0.48
The 2 nd pilot study							
No	Component	Item reliability	Separation	INFIT MNSQ		OUTFIT MNSQ	
				Max	Min	Max	Min
1	Vigor	.95	4.40	2.52	0.55	2.20	0.59
2	Initiative	.99	9.88	1.83	0.57	2.11	0.55
3	Emotional Closeness	1.00	15.64	1.28	0.70	1.37	0.69
4	Collaboration	.99	8.98	1.65	0.70	1.61	0.69
5	Adaptiveness	.99	8.33	1.87	0.48	2.09	0.49

*Note: N = 95 for the 1st pilot study (after cleaning 9 invalid responses from N = 104); N = 307 for the 2nd pilot study.

In the first stage of testing, after the data-cleaning procedure was carried out, 95 of 104 data points were analyzed further, and the reliability test results are presented in Table 1. The item reliability measurement indicated the degree to which items fit the Rasch measurement model. This means that the items used were reliable, and the number of items was sufficient to measure each construct. The accepted reliability index is $> .8$ (Bond & Fox, 2007). This indicates that the developed measuring instrument has good reliability. However, several items in the same construct had almost the same logit value. If these items were retained, they would have low discriminative power; therefore, the researcher decided to carry out reconstruction by revising (adding/reducing/changing) the words. After expert revision and inspection, the second testing stage, involving 307 teachers as participants, had a Pearson reliability value of .89. The reliability analysis results in the Phase 2 trial generally showed that the item reliability values for all constructs were very satisfactory ($> .95$).

Analysis of the Rasch model

The Rasch model concurrently depicts the interactions between respondents and items. Rather than relying on raw scores, this model is based on logit values, which represent the probability of a given group of respondents selecting a particular item. This approach is employed to project the raw score derived from Likert ratings, which are ordinal in nature (San Martín & Rolin, 2013). The psychometric measures used in this research included reliability at the instrument level (Table 1), item validity and item fit, instrument unidimensionality, item bias detection, and differential item functioning (DIF). Item validity in the initial stage refers to the polarity of test items. The point-measure correlation coefficient (PTMEA CORR) was used to determine the polarity of an item. According to the criteria recommended by Linacre (2009) and Ariffin et al. (2010), PTMEA CORR values should be positive, with values above .20 indicating that an item contributes meaningfully to measuring the intended construct. A value that is negative or equal to zero suggests that the response given to an item or by a respondent is inconsistent with the variable or construct being measured (Linacre, 2009). To identify invalid items, the mean-square (MNSQ) value was used in the next stage (Linacre, 2009). Linacre (2009) stated that items' MNSQ values should be between 0.5 to 1.5 (mean Infit MNSQ) (Sumintono & Widhiarso, 2013). Sumintono and Widhiarso (2013) described the MNSQ value as reflecting item accuracy, and should be between 0.4 to 1.4 for instruments with a Likert scale. According to Boone et al. (2014), the guidelines for assessing suitability criteria for test items are as follows:

- (1) Accepted Mean-Square (MNSQ) values: $0.5 < \text{MNSQ} < 1.5$
- (2) Accepted Z-standard outfit (ZSTD): $-2.0 < \text{ZSTD} < +2.0$

*If an item does not meet the criteria, it should be repaired or replaced (Boone et al., 2014).

It is important to consider the difficulty level of each question during the test. A useful tool for this involves generating a map that shows the distribution of subject abilities and difficulty levels of test items on the same scale. This map is called the Wright Map (Wright, 1977). The values on this map were arranged based on size values ranging from -3 to +3. However, in this case, a logit value above +2 or below -2 is considered an extreme value. Sumintono and Widhiarso (2015) offered an assessment guide consisting of four item categories based on their measurement values: 1) Very Easy (< -1), 2) Easy (-1 to 0), 3) Difficult (0 to 1), and 4) Very Difficult (> 1). Unidimensionality analysis uses Principal Component Analysis (PCA) of residuals, allowing measurement models (Baghaei, 2017). Nevertheless, Rasch recognized one factor that could pose a threat to unidimensionality. According to Baghaei and Aryadoust (2015), a value of unexplained variance in the first PCA construct residual of $< 3\%$ is considered special; 3–5% is very strong; 5–10% is firm; 10–15% is sufficient; and $> 15\%$ is weak.

Result

Findings

The Online Teacher Engagement Scale (OTES) was divided into five measured constructs: vigor, initiative, emotional closeness, collaboration, and adaptiveness, each consisting of eight items. The following are the results of the psychometric analysis of the OTES:

Item validity and item fit

Table 2 presents the outcomes of the validity and fit tests conducted on the items. Originally, the output generated by Winsteps was ordered according to the degree of difficulty, but the researcher edited and rearranged the questions in order. The measurement score served as an indicator of how difficult the test items were. The MNSQ fit values indicated whether the items fit the Rasch model. PTMEA CORR refers to the Point Measure Correlation, which is considered similar to the Point-Biserial Correlation in CTT. This parameter indicates the strength of the differentiating power.

Table 2.

Item Validity and Item Fit

Construct	Item	Score Measure (logit)	Model S.E.	Infit MNSQ	Outfit MNSQ	PTMEA CORR
Vigor	S1	0.45	0.09	0.92	1.11	.54
	S2	-0.65	0.11	2.52	2.20	.48
	S3	-0.09	0.10	0.90	0.97	.54
	S4	-0.17	0.11	0.57	0.59	.64
	S5	-0.39	0.11	0.93	0.99	.54
	S6	-0.41	0.11	0.76	0.74	.62
	S7	0.23	0.10	0.55	0.62	.62
	S8	1.03	0.07	1.16	1.60	.56
Initiative	I1	0.12	0.10	0.81	0.82	.56
	I2	-1.33	0.11	0.73	0.68	.63
	I3	0.41	0.10	0.92	0.99	.60
	I4	-0.77	0.11	0.74	0.70	.64
	I5	-1.07	0.11	0.57	0.55	.66
	I6	-0.54	0.11	0.61	0.59	.69
	I7	1.47	0.08	1.83	2.11	.32
	I8	1.70	0.08	1.34	1.71	.46
Emotional Closeness	KE1	-1.72	0.12	0.76	0.72	.52
	KE2	-1.86	0.12	1.28	1.31	.43
	KE3	-1.11	0.11	0.70	0.72	.53
	KE4	-0.66	0.10	0.74	0.88	.48
	KE5	1.74	0.08	1.27	1.34	.61
	KE6	2.11	0.07	0.99	1.09	.59
	KE7	-0.70	0.11	0.71	0.69	.53
	KE8	1.90	0.08	1.22	1.37	.56
Collaboration	Ko1	-1.53	0.12	0.88	0.83	.57
	Ko2	-0.86	0.12	1.65	1.61	.48
	Ko3	1.86	0.09	0.99	1.12	.62
	Ko4	-0.30	0.11	1.01	1.02	.54
	Ko5	1.01	0.10	0.95	1.00	.64
	Ko6	0.19	0.11	0.70	0.69	.69
	Ko7	0.42	0.11	0.83	0.78	.69
	Ko8	-0.80	0.12	1.03	1.08	.45
Adaptiveness	A1	-0.54	0.09	0.87	0.86	.51
	A2	0.59	0.08	1.35	1.45	.53

Table 2.
Item Validity and Item Fit

Construct	Item	Score Measure (logit)	Model S.E.	Infit MNSQ	Outfit MNSQ	PTMEA CORR
	A3	-0.75	0.10	0.59	0.59	.51
	A4	-0.44	0.09	0.61	0.62	.55
	A5	1.28	0.07	0.92	1.03	.55
	A6	0.69	0.08	1.87	2.09	.01
	A7	-1.17	0.10	0.48	0.49	.60
	A8	0.34	0.08	0.92	0.98	.53
Mean		0.00	0.09	0.98	1.07	
SD		0.96	0.01	0.46	0.56	

The validity of the test items refers to their polarity, as shown in Table 2. The polarity of each item was determined using the PTMEA CORR point-measure correlation coefficient. Findings indicate that item polarity within the vigor construct was between .48 to .64; the initiative construct was .50 to .60; the emotional closeness construct was .43 to .61; the collaboration construct was .45 to .69; and the adaptiveness construct was .01 to .60. Following the criteria proposed by Linacre (2009), the PTMEA CORR value should exceed .20, only one item in the adaptiveness construct, namely item A6, which had a PTMEA CORR index < .20, was removed from the OTES.

To determine how suitable and accurate the items were, item fit was assessed using the MNSQ value. In accordance with the criteria set out by Bond and Fox (2007) and Linacre (2009), the results showed that all the items fit the Rasch model. However, the unfeasible items must be removed. To calculate the maximum value, the average mean-square value (mean Infit MNSQ) (Sumintono & Widhiarso, 2013), which was 0.98, was added to the standard deviation (SD Infit MNSQ), which was 0.46, resulting in a value of 1.44. This meant that any items with Infit MNSQ values greater than 1.44 were considered not feasible. From Table 2, items S2, I7, Ko2, and A6 had Infit MNSQ values of more than 1.44, and were, therefore, removed.

As shown in Table 2, the mean item logit value was 0.0 logit, suggesting that the instrument overall could function as an effective measure. The average item logit value of 0.0 was a random value set to represent an equal 50:50 probability balance between the ability level of respondents and the difficulty level of the items (Bond & Fox, 2007). Overall, if the average logit of the items is not 0.0, the instrument is not effective. Upon further examination, no item within the same construct had the same logit value. This indicates that no item measures the same thing (Sumintono & Widhiarso, 2013). The distribution of respondents' ability levels and item difficulty can be observed in the Wright Map.

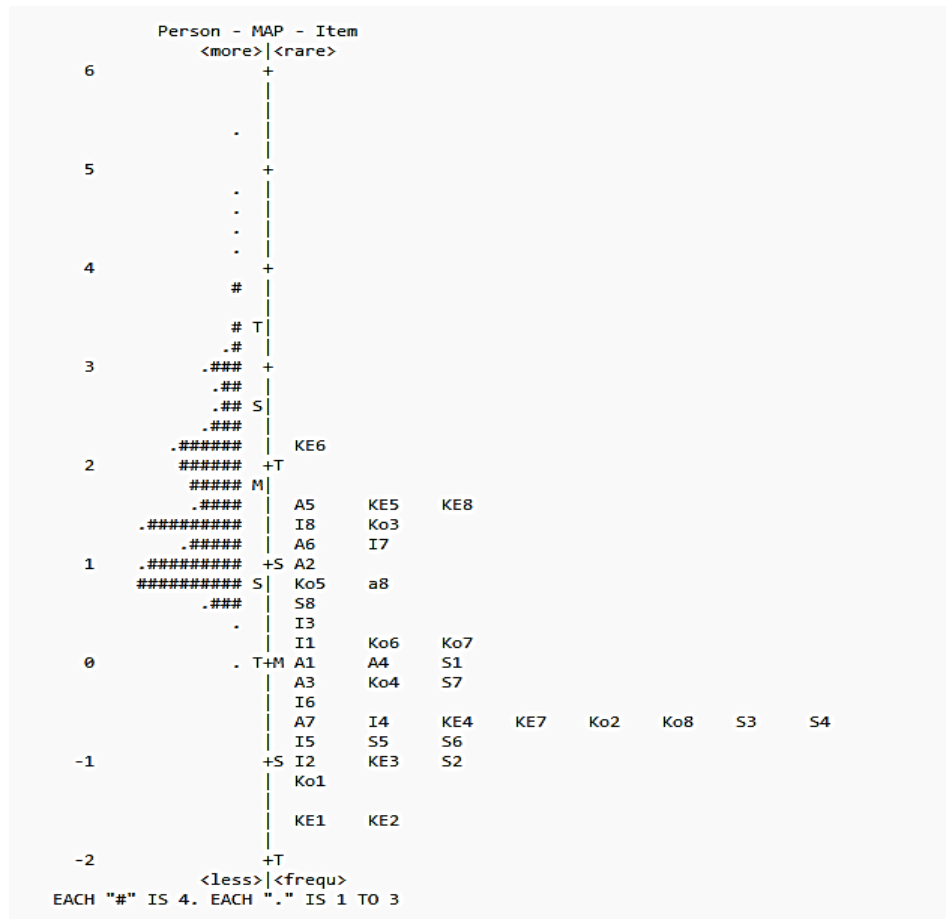


Figure 2. Person-Item Map

Figure 2 shows a person-item map of the data. The right-hand numbers refer to the items, and the left-hand numbers refer to the persons. Items at the top of the scale (above zero) were more difficult (to agree with). However, items at the bottom of the scale (below zero) were easier (to agree with). The lowest threshold was 1.86 logits (item KE2), and the highest was 2.11 logits (item KE6), meaning that, for most subjects, item KE2 was the most easily agreed upon, while item KE6 was the most difficult. These two items were ultimately able to differentiate between subjects who had the highest and lowest engagement in online learning. Four items (S2, I7, Ko2, and A6) were discarded because they had validity and item-fit values that did not meet the criteria.

Unidimensionality test

The concept of unidimensionality holds that every item in an instrument is intended to assess one common ability (Baghaei, 2017; Tabatabaee-Yazdi et al., 2018). This model provides a concise procedure for checking the instrument's dimensionality based on PCA (Baghaei & Aryadoust, 2015). Table 3 shows that the measuring tool was suitable for assessing vigor, initiative, emotional closeness, collaboration, and adaptiveness, which measures one dimension, namely, teacher engagement in online learning.

Table 3.

Unidimensionality Test

No.	Construct	Variance explained by measure (%)	Unexplained variance in 1 st contrast (%)
1.	Vigor	32.6	12.7
2.	Initiative	42.4	10.7

Table 3.*Unidimensionality Test*

No.	Construct	Variance explained by measure (%)	Unexplained variance in 1 st contrast (%)
3.	Emotional Closeness	60.1	11.0
4.	Collaboration	60.1	11.0
5.	Adaptiveness	34.0	16.6
	Whole Instrument	38.0	7.3

Table 3 indicates that the minimum requirement of 20% for unidimensionality was satisfied even though the vigor and adaptiveness constructs did not meet Rasch unidimensionality (> 40%) in the explained variance. suggesting that the instrument's level of item independence could be classified as good. In this sense, a dimension categorized as weak by a threshold such as > 15% is not necessarily invalid or worthless. Instead, it may reflect a more complex, partial contribution to the overall assessment of rationality or model performance (Siscoe, 2024). The unexplained variance of 16.6%, categorized as weak, warrants a more thorough discussion to clarify its implications for the construct of adaptiveness, especially in the context of teachers' digital skills and related competencies. Adaptiveness is inherently a complex, multifaceted construct influenced by numerous factors, including individual differences in digital literacy, teaching experience, and psychological attributes. Studies have repeatedly shown that digital literacy among teachers is uneven, with many educators reporting low self-perception of their digital skills and highlighting the urgent need for targeted professional development programs to address these gaps (Sánchez-Cruzado et al., 2021; Anggraini et al., 2024).

Differential item functioning (DIF) based on gender

Differential item functioning (DIF) tests whether the items are biased according to respondents' gender. A good item should be balanced. According to Sumintono and Widhiarso (2013), an item is categorized as gender-biased if the values for:

- (1) DIF contrast: $-0.5 < X < 0.5$
- (2) Rasch-Welch t: $-2.0 < X < 2.0$
- (3) Mantel-Haenszel Probability: $X < .05$

*Some authors have also used the chi-squared test to detect DIF (Lee et al., 2009).

Table 4.*Differential Item Functioning by Gender*

Item Number	DIF Contrast	Rasch-Welch t	Mantel-Haenszel Probability	Construct
S1	-0.68*	-3.05*	.0026*	Vigor
S2	0.61*	2.62*	.0086*	
S8	0.61*	3.42*	.0051*	
I2	-0.51*	-2.05*	.0442*	Initiative
A8	-0.65*	-3.31*	.0043*	Adaptiveness

Note: DIF contrast = the difference in the difficulty of the item between males and females. A negative DIF contrast indicates that the item is more difficult for females. *p < .05

Table 4 shows five items with DIF values outside the range specified in the three criteria. These items were: S1, S2, S8, I2, and A8. The five items that showed gender bias were removed from the OTES. These items indicate that there exists bias with gender as a differentiator in assessing teachers' level of engagement with online learning. The results in Table 4 were reinforced by Figure 3 below. For example, with item S1 of the vigor construct, men had a lower tendency to agree with it than women.

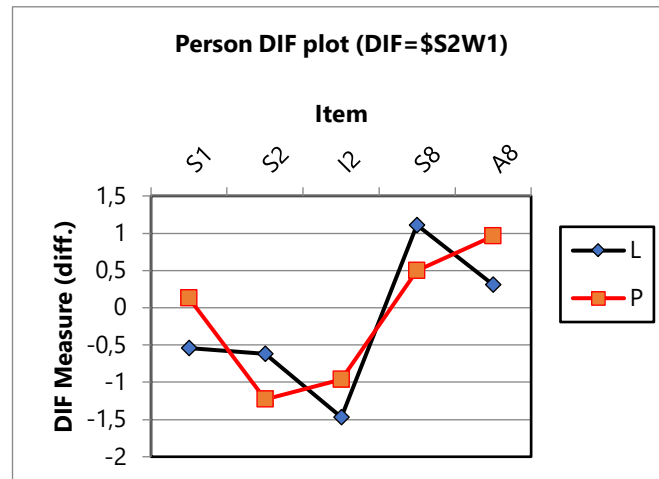


Figure 3. DIF Plot by Gender

After passing a series of item tests in the Rasch model, from the initial 40 compiled items, eight items were eliminated because they did not satisfy the requirements of the respective test conducted. The items that were discarded were S1, S2, S8, I2, I7, Ko2, A6, and A8. Thus, 32 items were valid for OTES.

Discussion

Item quality, consistency across items, and degree of differentiation, among other factors, can be assessed using Item Response Theory (IRT) analysis. This information is invaluable for improving the existing psychological tools. The five OTES constructs have good validity and reliability. Therefore, this scale can measure teacher engagement in online learning in terms of vigor, initiative, emotional closeness, collaboration, and adaptiveness. The vigor construct had an item reliability index of .95; for initiative, it was .99; emotional closeness was 1.00; collaboration was .99; and adaptiveness was .99. This means that the questionnaire was reliable (Linacre, 2009). The findings indicate that item polarity across all constructs ranged from .01 from .69. In line with the criteria proposed by Ariffin et al. (2010) and Linacre (2009), whereby the PTMEA CORR value should be greater than .20, only one item in the adaptiveness construct, namely item A6, which had a PTMEA CORR index < .20, was removed from the OTES. Only one item (A6) had a PTMEA CORR value of .01, which means that it was not as effective in measuring teacher engagement in online learning. Item quality, consistency, and differentiation within the five constructs of the Online Teacher Engagement Scale (OTES) were rigorously evaluated using Item Response Theory (IRT) analysis, which is a robust psychometric approach for assessing measurement properties of questionnaire items. The high item reliability indices reported .95 for vigor, .99 for initiative, 1.00 for emotional closeness, .99 for collaboration, and .99 for adaptiveness indicate that the scale items consistently measure their respective constructs with minimal measurement error. This strong reliability supports the scale's capacity to accurately capture various dimensions of teacher engagement in online learning environments, specifically reflecting teachers' energy, proactive behaviors, emotional connections, teamwork, and flexibility.

Beyond reliability, IRT enables the evaluation of both item discrimination, the extent to which an item distinguishes between respondents at different levels of the latent trait and item difficulty, indicating the trait level required for a higher likelihood of endorsing an item. This information is instrumental in refining scales by selecting items that are optimally informative across the engagement spectrum, ensuring precise measurement for all respondents (Reeve, 2003; Betz and Turner, 2011). Moreover, IRT allows for the detection of differential item functioning (DIF), whereby items performing inconsistently across subgroups can be identified, thereby enhancing fairness and validity (Zheng, 2021). These strengths in measurement precision and item-level diagnostics make IRT especially suitable for multidimensional constructs such as teacher engagement, which encompasses vigor, initiative, emotional closeness,

collaboration, and adaptiveness. The ability of IRT to anchor trait estimates to item content ensures that each construct is validly and reliably measured independently (Chen et al., 2018; Hays et al., 2000).

Additionally, the item polarity values ranging from .01 to .69, along with the PTMEA CORR (point-measure correlation) criteria suggest by Ariffin et al. (2010) and (Linacre (2009), further validate the scale's construct integrity. The PTMEA CORR values, which must be above .20, demonstrate that nearly all items contribute meaningfully to measuring teacher engagement. The exception of item A6, with a PTMEA CORR of .01, suggests it may not align well with the overall construct and could benefit from revision or removal to enhance the scale's precision. Overall, these psychometric findings affirm that the OTES is a reliable and valid instrument for assessing multifaceted teacher engagement in online learning contexts. The OTES specifically measures facets of teacher engagement that are critical in online teaching contexts, aligning with frameworks that highlight presence, expertise, engagement, and facilitation as key factors of online teaching effectiveness (Reyes-Fournier et al., 2020). The need for a teacher engagement scale in the online learning context remains critical due to several interconnected reasons grounded in the distinct challenges and dynamics of virtual education environments. First, teacher engagement significantly influences student achievement and motivation in online classrooms (Wang et al., 2022). Thus, measuring teacher engagement offers insight into a key driver of student success in online courses. Second, online learning environments inherently limit face-to-face teacher-student interactions, which can negatively affect the quality of instruction and student engagement. Research highlights the necessity to enhance teacher-student interactions in digital classrooms as limited engagement contributes to ineffective learning and increased dropout risk (Ong & Quek, 2023; Rabby et al., 2025). An accurate teacher engagement scale helps diagnose where engagement may be lacking and guides the improvement of teaching strategies suited for online platforms. Third, online learning comprises diverse modes synchronous and asynchronous that affect engagement differently. Studies identify multiple dimensions of engagement, including social, cognitive, behavioral, collaborative, and emotional components, which vary with instructional format (Yuyun, 2023). A teacher engagement scale tailored to online settings can capture these nuanced facets, helping evaluate how teachers foster engagement across various online instructional designs.

This study's major strength is its application of the Rasch model, which permits a level of critical psychometric analysis beyond what CTT alone can achieve. This study used an extensive dataset with strict data quality control mechanisms to develop a psychometrically sound instrument. With the implementation of online learning in Indonesia, the OTES scale could be an alternative instrument for measuring teacher engagement in online learning, analyzing problems in teacher training, and exploring teachers' enthusiasm, initiative, emotional closeness efforts, collaborative steps, and adaptiveness in carrying out online learning. The Online Teacher Engagement Scale (OTES), developed and validated using the Rasch model, demonstrates robust psychometric properties that underscore its effectiveness in measuring multifaceted teacher engagement within online learning contexts. The scale's five constructs vigor, initiative, emotional closeness, collaboration, and adaptiveness each exhibited high item reliability indices, ranging from 0.95 to 1.00, indicating strong internal consistency and precise measurement of these dimensions. The rigorous application of Item Response Theory (IRT) through Rasch analysis allowed for a nuanced examination of item quality, including differentiation and fit, which is critical for ensuring that each item meaningfully contributes to assessing teacher engagement. The removal of items that did not meet strict validity, fit, and differential item functioning (DIF) criteria further refined the scale, enhancing its accuracy and fairness across gender groups (Paek & Wilson, 2011; Misajon et al., 2016). This precision in measurement supports the OTES as a reliable instrument capable of capturing the complex behaviors and attitudes that characterize engaged teachers in digital learning environments.

Moreover, the study's methodological strengths lie in its use of the Rasch model, which overcomes limitations inherent in Classical Test Theory (CTT) by providing sample-independent item parameter estimates and accounting for varying measurement errors across respondents (Avinç & Doğan, 2024; Raykov & Marcoulides, 2011). This approach enables more accurate

assessment of latent traits such as engagement, which are often difficult to quantify. The OTES's validation process included comprehensive analyses such as unidimensionality testing and Wright mapping, confirming that the scale effectively measures a single underlying construct of teachers' engagement while differentiating among varying levels of respondent ability. Furthermore, the study highlights the necessity for additional validation using Classical Test Theory (CTT) methods to establish criterion-related validity. While preliminary analyses may support the instrument's reliability and construct validity, criterion-related validity assessment through CTT would provide a more robust evaluation of how well the test scores correlate with external criteria or outcomes. This step is crucial to confirm that the instrument accurately measures the intended construct and performs consistently across different populations and contexts. Future research should therefore focus on these validation processes to strengthen the instrument's psychometric properties and practical applicability (Bornstein, 2011).

Implication

The findings of this study provide theoretical, methodological, and practical implications for the development of instruments for measuring teacher involvement in online learning. Theoretically, the results of the study reinforce the conceptualization that teacher involvement is a multidimensional construct consisting of vigor, initiative, emotional closeness, collaboration, and adaptiveness, all of which have important contributions in explaining the effectiveness of online learning. The high reliability index in each dimension shows that the five constructs are able to consistently represent the main aspects of teacher engagement. These findings are in line with the concept of *teacher work engagement* which emphasizes that work energy, dedication, and adaptability are important determinants of teaching quality and student learning outcomes (Klassen et al., 2013; Schaufeli et al., 2002). Thus, OTES expands the literature on teacher engagement by providing instruments specifically designed for the context of online learning, which has different characteristics than face-to-face learning.

From a methodological perspective, this study shows that the use of the Rasch Model as part of Item Response Theory (IRT) provides advantages over the *Classical Test Theory* (CTT) approach in the development of psychometric instruments. Rasch's analysis not only evaluates reliability, but also identifies item quality through *item fit*, difficulty level, differentiating power, and *Differential Item Functioning* (DIF), resulting in a more objective instrument that is free from sample characteristic dependencies (Bond et al., 2021; Boone et al., 2014). The identification of A6 items that have low PTMEA CORR values indicates the importance of continuous evaluation of items so that the instrument remains constructively valid and has high measurement precision. Therefore, this study provides empirical evidence that the application of the Rasch approach is worthy of being a standard in the development of educational instruments, especially for measuring multidimensional psychological constructs.

Practically, OTES has broad potential as an evaluation instrument for schools, universities, and teacher training institutions in monitoring the level of teacher involvement during online learning. The information generated from the measurement of each dimension allows policymakers to identify areas that need intervention, such as increasing teachers' initiative in utilizing technology, strengthening emotional closeness with students, increasing professional collaboration, and developing adaptability to digital learning changes. The use of valid and reliable instruments also allows the evaluation of teacher training programs to be carried out more accurately so that professional development policies can be designed based on empirical evidence (Darling-Hammond et al., 2017).

In addition, this study's results have practical implications for improving the quality of online learning. Various studies show that teacher involvement is closely related to learning motivation, student engagement, learning satisfaction, and academic achievement in a digital learning environment (Wang et al., 2022; Ong & Quek, 2023). Therefore, OTES can be used as a diagnostic instrument to identify factors that inhibit teacher involvement before impacting the quality of the learning process and increasing the risk of *student disengagement* and *dropout*. Thus, the results of OTES measurements can be the basis for developing strategies to improve the quality of online learning that are more effective and oriented to the needs of teachers.

Nonetheless, the implications of this research also lead to the need for further development. Instrument validation through the CTT approach, *criterion-related validity testing*, and the application of OTES at various levels of education and different cultural contexts will further strengthen the generalization of the instrument. Cross-country and cross-platform testing of digital learning is also needed to ensure that the psychometric characteristics of OTES remain stable in various online learning environments. This step will increase the practical and academic value of OTES as a standard instrument in research and evaluation of teacher involvement in the digital education era (Bornstein, 2011; Raykov & Marcoulides, 2011).

Limitation

Overall, the OTES offers a psychometrically sound tool that can inform teacher training, support online learning implementation, and guide further cross-national research on teacher engagement in the evolving educational landscape. Despite the good psychometric values, this study had several limitations. First, as a self-report instrument, OTES is susceptible to social desirability bias, given that respondents may prefer to present a positive rather than negative depiction of their work. Second, although IRT provides indices of validity and reliability at the item and person levels, construct validity and criterion-related validity (McCreary et al., 2013; Stump et al., 2012) using CTT are also needed, especially regarding the effectiveness of OTES in predicting teacher engagement in online learning.

Conclusion

This study successfully developed and validated the Online Teacher Engagement Scale (OTES) using the Rasch measurement model, resulting in a psychometrically sound instrument for assessing teacher engagement in online learning environments. The final OTES demonstrated strong measurement properties, including high reliability, satisfactory item fit, appropriate construct validity, and the absence of gender bias, indicating that the instrument provides stable, fair, and accurate measurement across respondents. These findings confirm that the five dimensions of teacher engagement including vigor, initiative, emotional closeness, collaboration, and adaptiveness are represented by well-functioning items capable of capturing the multidimensional nature of teacher engagement in digital learning contexts.

The validated OTES with 32 items contributes to the growing need for context-specific instruments that can accurately evaluate teachers' engagement in online education. Beyond its psychometric strengths, the instrument offers practical value for educational institutions, policymakers, and researchers. It can be used to identify strengths and areas for improvement in teachers' engagement, evaluate professional development programs, support evidence-based decision-making, and inform interventions aimed at enhancing the quality of online teaching and learning. Consequently, OTES provides a valuable tool for both educational evaluation and future research on teacher engagement in technology-mediated learning environments.

Although the Rasch model provided robust evidence of the instrument's validity and reliability, further studies are recommended to strengthen its generalizability by examining the OTES across more diverse educational settings, larger and more heterogeneous samples, and different cultural contexts. Future research should also complement Rasch analysis with Classical Test Theory approaches, including criterion-related validity and confirmatory factor analysis, to provide additional evidence of the instrument's psychometric quality and broaden its applicability in international educational research.

Credit/Acknowledgement

The authors would like to express their sincere gratitude to all the teachers who generously participated in this study and shared their valuable time and insights during the data collection process.

Author Contributions

IFK contributed to conceptualization, methodology, formal analysis, and writing – original draft preparation. BS contributed to conceptualization and formal analysis. ES contributed to data curation and writing – review and editing. UP contributed to writing – review and editing. CNS contributed to writing – review and editing and manuscript formatting. All authors have read and agreed to the published version of the manuscript.

References

- Agasisti, T., Antequera, G., & Delprato, M. (2023). Technological resources, ICT use and schools efficiency in Latin America – Insights from OECD PISA 2018. *International Journal of Educational Development*, 99, 1-11. <https://doi.org/10.1016/j.ijedudev.2023.102757>
- Anastasopoulou, E., Tsagri, A., Avramidi, E., Lourida, K., Mitroyanni, E., Tsogka, D., & Katsikis, I. (2024). The impact of ICT on education. *Technium Soc. Sci. J.*, 58, 48-55. <https://access.heinonline.com/HOL/P?h=hein.journals/techssj58&i=49>
- Anggraini, T., Ahmad, M., & Hanafi, I. (2024). Digital literacy and teaching experience as predictors of pedagogical competence in the digital era. *Tarbawi: Jurnal Keilmuan Manajemen Pendidikan*, 10(02), 295-306. <https://doi.org/10.32678/tarbawi.v10i02.10795>
- Anita, A. (2024). E-Learning initiatives in teacher education. *Shodh Sari-An International Multidisciplinary Journal*, 03(01), 186–195. <https://doi.org/10.59231/SARI7664>
- Archambault, L., Leary, H., & Rice, K. (2022). Pillars of online pedagogy: A framework for teaching in online learning environments. *Educational Psychologist*, 57(3), 178–191. <https://doi.org/10.1080/00461520.2022.2051513>
- Ariffin, S. R., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability multiple intelligent item using rasch measurement model. *Procedia - Social and Behavioral Sciences*, 9, 729–733. <https://doi.org/10.1016/j.sbspro.2010.12.225>
- Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*, 29(17), 22895–22941. <https://doi.org/10.1007/s10639-024-12662-7>
- Baena-Navarro, R., Serrano-Ardila, L., & Carriazo-Regino, Y. (2024). Innovative model for the integration of ICTs in rural environmental education: Towards a sustainable pedagogy. *Southern Perspective/ Perspectiva Austral*, 2, 1-14. <https://doi.org/10.56294/pa202435>
- Baghaei, P. (2017). Application of multidimensional Rasch models in large scale assessment and validation: An empirical example. *Electronic Journal of Research in Education Psychology*, 10(26), 233–252. <https://doi.org/10.25115/ejrep.v10i26.1493>
- Baghaei, P., & Aryadoust, V. (2015). Modeling local item dependence due to common test format with a multidimensional Rasch model. *International Journal of Testing*, 15(1), 71–87. <https://doi.org/10.1080/15305058.2014.941108>
- Bakker, A. B., & Bal, M. P. (2010). Weekly work engagement and performance: A study among starting teachers. *Journal of Occupational and Organizational Psychology*, 83(1), 189–206. <https://doi.org/10.1348/096317909X402596>
- Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and work engagement: The JD–R approach. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 389–411. <https://doi.org/10.1146/annurev-orgpsych-031413-091235>
- Betz, N. E., & Turner, B. M. (2011). Using item response theory and adaptive testing in online career assessment. *Journal of Career Assessment*, 19(3), 274–286. <https://doi.org/10.1177/1069072710395534>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates Publishers.
- Bond, T. G., & Fox, C. M. (2012). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Routledge.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (4th ed.). Routledge.

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences* (Vol. 10, pp. 978-94). Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23(2), 532–544. <https://doi.org/10.1037/a0022402>
- Bradley, K. D., Peabody, M. R., Akers, K. S., & Knutson, N. M. (2015). Rating scales in survey research: Using the Rasch model to illustrate the middle category measurement flaw. *Survey Practice*, 8(1), 1–12. <https://doi.org/10.29115/SP-2015-0001>
- Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika*, 83(3), 538–562. <https://doi.org/10.1007/s11336-018-9610-4>
- Corry, M., & Stella, J. (2018). Teacher self-efficacy in online education: A review of the literature. *Research in Learning Technology*, 26, 1-12. <https://doi.org/10.25304/rlt.v26.2047>
- Cueva, A., & Inga, E. (2022). Information and communication technologies for education considering the flipped learning model. *Education Sciences*, 12(3), 1-16. <https://doi.org/10.3390/educsci12030207>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. *Learning policy institute*.
- Das, U. (2023). Online learning: Challenges and solutions for learners and teachers. *Management and Labour Studies*, 48(2), 210–213. <https://doi.org/10.1177/0258042X211069501>
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
- Fischer, H. F., & Rose, M. (2016). www.common-metrics.org: A web application to estimate scores from different patient-reported outcome measures on a common scale. *BMC Medical Research Methodology*, 16(1), 1-5. <https://doi.org/10.1186/s12874-016-0241-0>
- Grammens, M., Voet, M., Vanderlinde, R., Declercq, L., & De Wever, B. (2022). A systematic review of teacher roles and competences for teaching synchronously online through videoconferencing technology. *Educational Research Review*, 37, 1-16. <https://doi.org/10.1016/j.edurev.2022.100461>
- Greenier, V., Derakhshan, A., & Fathi, J. (2021). Emotion regulation and psychological well-being in teacher work engagement: A case of British and Iranian English language teachers. *System*, 97, 1-18. <https://doi.org/10.1016/j.system.2020.102446>
- Grönlund, A., Samuelsson, J., & Samuelsson, J. (2023). When documentation becomes feedback: Tensions in feedback activity in Learning Management Systems. *Education Inquiry*, 14(2), 194–212. <https://doi.org/10.1080/20004508.2021.1980973>
- Gudmundsdottir, Dr. G. B., & Hatlevik, O. E. (2020). “I just Google it”—Developing professional digital competence and preparing student teachers to exercise responsible ICT use. *Nordic Journal of Comparative and International Education (NJCIE)*, 4(3–4), 39–55. <https://doi.org/10.7577/njcie.3752>
- Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology*, 43(6), 495–513. <https://doi.org/10.1016/j.jsp.2005.11.001>
- Hakanen, J. J., Ropponen, A., Schaufeli, W. B., & De Witte, H. (2019). Who is engaged at work?: A large-scale study in 30 European countries. *Journal of Occupational & Environmental Medicine*, 61(5), 373–381. <https://doi.org/10.1097/JOM.0000000000001528>
- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical care*, 38(9), II-42. <https://doi.org/10.1097/00005650-200009002-00007>
- Ho, S. K., Chan, H., Chan, E. S.-S., Fung, K.-K., & Hung, S. S.-L. (2021). A preliminary psychometric investigation of a Chinese version of the Engaged Teachers Scale (C-ETS). *Frontiers in Education*, 6, 1-8. <https://doi.org/10.3389/educ.2021.676835>

- Wal, J. J. I. D., Brok, P. J., Hooijer, J. G., & Martens, R. L. (2014). Teachers' engagement in professional learning: Exploring motivational profiles. *Learning and Individual Differences*, 36, 27–36. <https://doi.org/10.1016/j.lindif.2014.08.001>
- Klassen, R. M., Yerdelen, S., & Durksen, T. L. (2013). Measuring teacher engagement: Development of the Engaged Teachers Scale (ETS). *Frontline Learning Research*, 1(2), 33–52. <https://doi.org/10.14786/flr.v1i2.44>
- Lee, Y. S., Cohen, A., & Toro, M. (2009). Examining type I error and power for detection of differential item and testlet functioning. *Asia Pacific Education Review*, 10(3), 365–375. <https://doi.org/10.1007/s12564-009-9039-7>
- Leite, W. L., Kuang, H., Jing, Z., Xing, W., Cavanaugh, C., & Huggins-Manley, A. C. (2022). The relationship between self-regulated student use of a virtual learning environment for algebra and student achievement: An examination of the role of teacher orchestration. *Computers & Education*, 191, 1–15. <https://doi.org/10.1016/j.compedu.2022.104615>
- Li, M., Wang, Z., You, X., & Gao, J. (2015). Value congruence and teachers' work engagement: The mediating role of autonomous and controlled motivation. *Personality and Individual Differences*, 80, 113–118. <https://doi.org/10.1016/j.paid.2015.02.021>
- Liang, R., & Chen, D.-T. V. (2012). Online learning: Trends, potential and challenges. *Creative Education*, 3(8), 1332–1335. <https://doi.org/10.4236/ce.2012.38195>
- Lin, R., Yang, J., Jiang, F., & Li, J. (2023). Does teacher's data literacy and digital teaching competence influence empowering students in the classroom? Evidence from China. *Education and Information Technologies*, 28(3), 2845–2867. <https://doi.org/10.1007/s10639-022-11274-3>
- Linacre, J. M. (2009). Local independence and residual covariance: A study of Olympic figure skating ratings. *Journal of Applied Measurement*, 10(2), 157–169. <https://pubmed.ncbi.nlm.nih.gov/19564696/>
- McCreary, L. L., Conrad, K. M., Conrad, K. J., Scott, C. K., Funk, R. R., & Dennis, M. L. (2013). Using the Rasch measurement model in psychometric analysis of the family effectiveness measure. *Nursing Research*, 62(3), 149–159. <https://doi.org/10.1097/NNR.0b013e31828eafe6>
- Misajon, R., Pallant, J., & Bliuc, A.-M. (2016). Rasch analysis of the personal wellbeing index. *Quality of Life Research*, 25(10), 2565–2569. <https://doi.org/10.1007/s11136-016-1302-x>
- Ngodu, A., Ndibalema, P. M., & William, F. (2024). Context-relevant strategies for ICT integration in teaching and learning science subjects in Tanzania secondary schools. *Educational Technology Quarterly*, 2024(1), 20–37. <https://doi.org/10.55056/etq.704>
- Ong, S. G. T., & Quek, G. C. L. (2023). Enhancing teacher–student interactions and student online engagement in an online learning environment. *Learning Environments Research*, 17(1), 681–707. <https://doi.org/10.1007/s10984-022-09447-5>
- Osterlind, J. (2009). *Modern measurement: Theory, principles, and applications of mental appraisal*. Pearson.
- Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel–Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement*, 71(6), 1023–1046. <https://doi.org/10.1177/0013164411400734>
- Penttinen, V., Pakarinen, E., Von Suchodoletz, A., & Lerkkanen, M.-K. (2020). Relations between kindergarten teachers' occupational well-being and the quality of teacher-child interactions. *Early Education and Development*, 31(7), 994–1010. <https://doi.org/10.1080/10409289.2020.1785265>
- Perera, H. N., Vosicka, L., Granziera, H., & McIlveen, P. (2018). Towards an integrative perspective on the structure of teacher work engagement. *Journal of Vocational Behavior*, 108, 28–41. <https://doi.org/10.1016/j.jvb.2018.05.006>
- Rabby, S. M., Akhi, A. A., Talukder, M. S., & Das, S. (2025). How does limited student engagement in virtual classrooms affect students' academic performance? A stressor–strain–outcome

- perspective. *On the Horizon: The International Journal of Learning Futures*, 33(3), 357–377. <https://doi.org/10.1108/oth-02-2025-0017>
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge. <https://doi.org/10.4324/9780203841624>
- Reeve, B. B. (2003). Item response theory modeling in health outcomes measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*, 3(2), 131–145. <https://doi.org/10.1586/14737167.3.2.131>
- Reyes-Fournier, E., Cumella, E. J., March, M., Pedersen, J., & Blackman, G. (2020). Development and validation of the Purdue Global Online Teaching Effectiveness Scale. *Online Learning*, 24(2), 111–127. <https://doi.org/10.24059/olj.v24i2.2071>
- Richardson, J. C., Besser, E., Koehler, A., Lim, J., & Strait, M. (2016). Instructors' perceptions of instructor presence in online learning environments. *International Review of Research in Open and Distributed Learning*, 17(4), 82–104. <https://doi.org/10.19173/irrodl.v17i4.2330>
- Sánchez-Cruzado, C., Santiago Campión, R., & Sánchez-Compañía, M. T. (2021). Teacher digital literacy: The indisputable challenge after COVID-19. *Sustainability*, 13(4), 1–29. <https://doi.org/10.3390/su13041858>
- San Martín, E., & Rolin, J.-M. (2013). Identification of parametric Rasch-type models. *Journal of Statistical Planning and Inference*, 143(1), 116–130. <https://doi.org/10.1016/j.jspi.2012.06.014>
- Sarosa, S. (2022). The effect of perceived risks and perceived cost on using online learning by high school students. *Procedia Computer Science*, 197, 477–483. <https://doi.org/10.1016/j.procs.2021.12.164>
- Schaufeli, W. B., Salanova, M., González-romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3(1), 71–92. <https://doi.org/10.1023/A:1015630930326>
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and psychological measurement*, 66(4), 701–716. <https://doi.org/10.1177/0013164405282>
- Sidi, Y., Shamir-Inbal, T., & Eshet-Alkalai, Y. (2023). From face-to-face to online: Teachers' perceived experiences in online distance teaching during the Covid-19 pandemic. *Computers & Education*, 201, 104831. <https://doi.org/10.1016/j.compedu.2023.104831>
- da Silva Júnior, D. I., Ferreira, M. C., & Valentini, F. (2020). Psychometric properties of the engaged teachers scale (ETS). *Psicologia: Teoria e prática*, 22(1), 109–126. <https://doi.org/10.5935/1980-6906/psicologia.v22n1p109-126>
- Siscoe, R. W. (2024). Justification as a dimension of rationality. *Philosophical Studies*, 181(6), 1523–1546. <https://doi.org/10.1007/s11098-024-02131-9>
- Soininen, V., Pakarinen, E., & Lerkkanen, M.-K. (2023). Reciprocal associations among teacher–child interactions, teachers' work engagement, and children's social competence. *Journal of Applied Developmental Psychology*, 85, 1–13. <https://doi.org/10.1016/j.appdev.2022.101508>
- Stump, G. S., Husman, J., & Brem, S. K. (2012). The nursing student self-efficacy scale: Development using item response theory. *Nursing Research*, 61(3), 149–158. <https://doi.org/10.1097/NNR.0b013e318253a750>
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. TrimKom Publishing House.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Penerbit Trim Komunikata.
- Tabatabaee-Yazdi, M., Motalebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and validation of a teacher success questionnaire using the Rasch model. *International Journal of Instruction*, 11(2), 129–144. <https://doi.org/10.12973/iji.2018.11210a>
- Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: Early developments, current practice, and future

- challenges. *Frontiers in Rehabilitation Sciences*, 4, 1-17. <https://doi.org/10.3389/fresc.2023.1208670>
- Turnbull, D., Chugh, R., & Luck, J. (2020). Learning management systems, an overview. In A. Tatnall (Ed.), *Encyclopedia of education and information technologies*. Springer. <https://doi.org/10.1007/978-3-030-10576-1>
- Wang, J., Zhang, X., & Zhang, L. J. (2022). Effects of teacher engagement on students' achievement in an online English as a foreign language classroom: The mediating role of autonomous motivation and positive emotions. *Frontiers in Psychology*, 13, 1-11. <https://doi.org/10.3389/fpsyg.2022.950652>
- Wasson, B., & Hansen, C. (2014). Making use of ICT: Glimpses from Norwegian teacher practices. *Nordic Journal of Digital Literacy*, 9(1), 44–65. <https://doi.org/10.18261/ISSN1891-943X-2014-01-06>
- Wilson, M. (2023). *Constructing measures: An item response modeling approach* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003286929>
- Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14(2), 97–116. <https://doi.org/10.1111/j.1745-3984.1977.tb00031.x>
- Yerdelen, S., Durksen, T., & Klassen, R. M. (2018). An international validation of the engaged teacher scale. *Teachers and Teaching*, 24(6), 673–689. <https://doi.org/10.1080/13540602.2018.1457024>
- Yuyun, I. (2023). Investigating university student engagement in online learning: A case study in EFL classroom. *Indonesian Journal of Applied Linguistics*, 12(3), 648–667. <https://doi.org/10.17509/ijal.v12i3.46035>
- Zheng, X., & Yang, J. S. (2021). Multiple group Item Response Theory applications using Stata IRT Package. *Measurement: Interdisciplinary Research and Perspectives*, 19(3), 190-198. <https://doi.org/10.1080/15366367.2021.1911507>