European Proceedings of
Social and Behavioural Sciences
EpSBS

www.europeanproceedings.com
e-ISSN: 2357-1330

DOI: 10.15405/epsbs.2026.09.77

ICMC 2025
The 4th International Conference on Management and Communication

PSYCHOMETRIC PROPERTIES OF AN INSTRUMENT MEASURING TEACHERS' COMPETENCY IN CONDUCTING STORYTELLING

Norazly Nordin (a)*, Joanne Lo Fui Senn (b), Muhammad Hilmi Abdul Rashid (c)
*Corresponding author

(a) Tunku Abdul Rahman University of Management and Technology, Sabah Branch, Malaysia, norazly@tar.edu.my
(b) Tunku Abdul Rahman University of Management and Technology, Sabah Branch, Malaysia
(c) Tunku Abdul Rahman University of Management and Technology, Pahang Branch, Malaysia

Abstract

This study investigates the psychometric properties of an instrument designed to assess the competency levels of preschool teachers in Malaysia in conducting storytelling. A sample of 90 preschool teachers participated in the study. The Rasch measurement model was employed to evaluate the reliability and validity of the instrument, focusing on respondent and item reliability indices, item polarity, and item fitness. The results indicated a high respondent reliability index of 0.94 and a separation index of 3.86. The item reliability was also satisfactory, with an index of 0.81 and a separation index of 2.03. All items demonstrated positive PTMEA CORR values, confirming appropriate item polarity. However, the item fitness analysis revealed that five items were misfit and required revision, while 19 items were deemed acceptable. These findings suggest that the instrument is generally reliable and valid for assessing teachers' competency in storytelling, though some items need refinement. The study highlights the importance of continuous evaluation and improvement of assessment tools to ensure accurate measurement of educational competencies.

Keywords: Rasch measurement model, competency, storytelling, validity and reliability

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

Storytelling is one of the oldest forms of communication and is used to share experiences, understand others, and entertain. Storytelling is also an age-old tradition passed down through generations, serving as a powerful tool for education, especially in early childhood education. For young children, storytelling is a source of amusement and a vital component of their cognitive, emotional, and social development. Through stories, children are transported to imaginative worlds to explore new ideas, learn valuable life lessons, and develop essential language skills. Using storytelling as a pedagogical tool in early childhood education provides many advantages in children's development, particularly in character building; language and vocabulary development; cognitive and critical thinking; and social and emotional development (Bidari, 2019; Burdelski & Evaldsson, 2019; Otoluwa, Talib et al., 2022). By integrating storytelling into early education, educators can effectively nurture well-rounded individuals. For this reason, storytelling is strongly emphasized in teaching and learning in early childhood education.

Storytelling to young children requires careful planning and engaging techniques. Teachers need to have a set of skills to ensure that storytelling is conducted effectively. According to Isbell and Raines (2013), skills that teachers need to master in handling storytelling sessions are selecting an age-appropriate story, familiarizing the characters and plot, and questioning techniques. Miller and Pennycuff (2008) also added teachers need to be competent in engaging children by asking predictive or reflective questions and able to incorporate props or puppets to enhance interaction and comprehension, maintain eye contact, and encourage participation, such as repeating phrases or mimicking actions. Teachers need to be competent in storytelling to young children as it is a powerful educational tool that promotes language development, imagination, and social-emotional learning.

Despite the recognised benefits of storytelling, there is still limited understanding of how competent early childhood educators are in applying storytelling strategies in real classroom settings. Most existing research focuses on children's developmental gains rather than teachers' mastery of storytelling skills, the challenges they encounter, or how their competencies influence children's engagement and learning. In practice, competent storytelling enables teachers to capture attention, adapt delivery to children's developmental levels, and foster emotional connection, empathy, and problem-solving. Skilled storytellers also enhance classroom interaction through techniques such as role-playing and open-ended questioning, helping build participation, critical thinking, and a positive teacher–child relationship.

1.1. Research Objectives

The objectives of this study are:

i. To investigate the reliability of the survey Instrument of Teacher's Competency Level in Conducting Storytelling.
ii. To investigate the validity of the survey Instrument of Teacher's Competency Level in Conducting Storytelling.

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

2. Literature Review

A competency instrument is a tool designed to measure or assess an individual's competencies, which are the knowledge, skills, abilities, and behaviors required to perform a specific task, role, or function effectively. These instruments are often used in educational, organizational, and professional settings to evaluate proficiency levels, identify training needs, and support decision-making in recruitment, performance appraisal, and career development (Filomeno et al., 2024; Yaqoob Mohammed Al Jabri et al., 2021; Rusli et al., 2024).

In the education sector, a competencies instrument is a tool designed to assess and measure the skills and abilities that are deemed essential for students or educators. These instruments are crucial for evaluating competencies such as digital literacy, communication, collaboration, critical thinking, and other key skills necessary for modern educational and professional environments. Common methods for assessing competencies include surveys, questionnaires, and rubrics. These tools are used to measure both digital competencies and soft skills, ensuring that students and educators are able to achieve the desired outcomes Cruz et al., 2019; Peláez-Sánchez et al., 2024). Few studies have been carried out on the development of the competency instrument specifically to assess the teacher's competencies in various aspect such as assessing pedagogical skills (Ludwikowska, 2022), technical and vocational skills (Arifin et al., 2021), digital competencies (Ghomi & Redecker, 2019), subject-specific competency instruments like science and technology (Eliyawati et al., 2023), blended teaching competency instrument (Archibald et al., 2020), and language teacher competency instrument (Montenegro-Rueda, & Fernández-Batanero, 2023).

A good competency instrument must demonstrate strong validity and reliability. Validating and checking the reliability of competency instruments is crucial to ensure they accurately measure what they are intended to and produce consistent results across different contexts and populations. This process is essential for the credibility and utility of the instruments in assessing competencies in various fields. According to Heale and Twycross (2015), validity refers to the extent to which an instrument measures what it is intended to measure. It ensures that the instrument accurately reflects the concept it is supposed to assess and reliability refers to the consistency of an instrument in measuring a concept. It indicates the degree to which the instrument yields the same results under consistent conditions. In other words, validity ensures that an instrument measures the intended concept accurately, while reliability ensures that the measurement is consistent across different instances. Both are crucial for the credibility and utility of competency instruments, as they help minimize errors and enhance the quality of measurement.

According to Vetter and Cubbin, (2019), there are three types of validity that need to be highlighted in assessing the instrument which are content, construct, and criterion validity. Content validity refers to the extent to which a measurement instrument, such as a test or survey, accurately represents all aspects of the construct it aims to measure. This involves ensuring that the instrument covers all relevant parts of the concept and excludes any irrelevant elements. Meanwhile, construct validity refers to the extent to which a test or measurement tool accurately measures the theoretical construct it is intended to measure. This involves ensuring that the test truly reflects the concept it aims to assess, rather than something else and Criterion validity checks if a test or measurement can accurately predict an outcome by comparing it to an already established and trusted measure. Cooper et al. (2020) suggested the method of assessing the validity of the instrument as explained in Table 1 below:

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

Table 1. Method in assessing validity of the instrument

Type of Validity	Method	Description
Criterion	Concurrent Validity	Compares the test results with an established measure taken at the same time to see if they correlate.
	Predictive Validity	Evaluates how well the test predicts future outcomes by comparing test results with future performance.
Construct	Convergent Validity	Assesses whether the test correlates well with other tests measuring the same construct.
	Discriminant Validity	Ensures the test does not correlate with tests measuring different constructs.
Content	Expert Judgment	Involves subject matter experts evaluating whether the test items represent the construct comprehensively.
	Content Validity Index (CVI)	Quantifies content validity by calculating the proportion of items rated as relevant by experts.

For the reliability process, the common method is using Cronbach's alpha to measure the consistency of results across items within the instrument. Cronbach's alpha quantifies how well a set of items measures a single, unidimensional construct. It ranges from 0 to 1, with higher values indicating greater reliability (Cooper et al., 2020). Vetter and Cubbin (2019) also explained that reliability can be assessed through a test-retest approach. This method evaluates how stable and consistent the results of a test or measurement are when administered to the same group of people at two different points in time. The results are typically analyzed using a statistical correlation coefficient, such as Pearson's r or Spearman's ρ depending on the data type. A high positive correlation (close to +1) indicates good reliability. In conclusion, assessing the validity and reliability of psychometric tools involves a combination of methods to ensure comprehensive evaluation. Content, construct, and criterion validity are key for establishing what the tool measures, while internal consistency and test-retest reliability ensure consistent measurement. These assessments are essential for the development and application of robust psychometric instruments.

One of the approaches in assessing the psychometric properties of the instrument is using the Rasch Measurement Model. Rasch Model is particularly useful for measuring latent traits that cannot be directly observed, such as cognitive abilities or psychological attributes. The model estimates both the person's ability and the item's difficulty, and the difference between these determines the probability of a correct response (Tesio et al., 2023). Besides, this model is widely used in various fields, including language assessment, rehabilitation medicine, and educational testing, to ensure that measurements are valid and reliable (Aryadoust et al., 2020; Boone & Staver, 2020; Nordin et al., 2023; Wang, 2024).

Bond and Fox (2013) stated that there have the criteria of validity and reliability need to be fulfilled for the instrument according to the Rasch Measurement model. The criteria are explained in Table 2 below:

Table 2. The Criteria of validity and reliability according to the rasch model

Criteria	**Statistic Data**	**Minimum Requirement**
Validity	Item Polarity	PTMEA CORR > 0
	Item Fit	Total Mean Square (MNSQ) infit and outfit of 0.6 – 1.4, and Z-Standard (ZSTD) between the range -2 to 2.
Reliability	Person Reliability	Value > 0.8

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

	Item Reliability	Value > 0.8
	Cronbach Alpha	Value > 0.7
	Separation index	Value > 2.00

The Rasch Measurement Model is a powerful tool for assessing the validity and reliability of research instruments. Its ability to transform data, ensure unidimensionality, and provide detailed reliability and validity measures makes it a preferred choice over traditional methods like Classical Test Theory. This model enhances the precision and interpretability of research findings, making it invaluable in educational and psychological assessments (Aryadoust & Sayama, 2020; Fischer et al., 2021; Tesio et al., 2023). Assessing the validity and reliability of an instrument is crucial for ensuring the accuracy and consistency of the measurements it provides. Validity ensures that the instrument truly measures what it is intended to measure, covering all relevant aspects of the construct. Reliability, on the other hand, ensures that the instrument produces stable and consistent results over time and across different conditions. By rigorously evaluating both validity and reliability, researchers and practitioners can trust that their instruments yield meaningful and dependable data. This, in turn, enhances the credibility of their findings, supports sound decision-making, and contributes to the overall quality of research and practice. In summary, assessing validity and reliability is essential for developing effective, trustworthy instruments that can accurately capture the constructs of interest and provide reliable insights.

3. Methodology

This study utilized a quantitative research design to investigate the validity and reliability of a newly developed survey instrument aimed at measuring teachers' competency level in storytelling. Surveys provide a standardized method of data collection, which is essential for assessing psychometric properties like reliability and validity. Standardization ensures that all participants respond to the same set of questions under similar conditions, making the results more comparable and reliable. Surveys allow researchers to gather data from a large and diverse sample of teachers, providing a broad understanding of their competencies. This helps ensure that the instrument is tested across various contexts and teaching environments Based on a comprehensive review of relevant literature, the survey consisted of 24 items in three domains. The distribution of items in the domain and items in each domain is shown in Table 3 and Table 4 as follows:

Table 3. Distribution of items in the domains

Domain	Before Storytelling	During Storytelling	After Storytelling
No of Item	8 items	10 items	6 items

Table 4. Items in each domain

Domain	Item	Code
Before story telling	I am able to choose appropriate story for preschool children	A1
	I am able to conduct my own rehearsal before conducting storytelling to children.	A2

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

	I am able to hook children attention to listen my story	A3
	I am able to set up the area of storytelling to lighten the mood of the children.	A4
	I am able to produce the appropriate props for my storytelling session.	A5
	I am able to understand the story before telling it to the children.	A6
	I am able to create vocal sound effects to create a more immersive experience in my storytelling session.	A7
	I am able to plan my storytelling according to the time given.	A8
During storytelling	I am able to use variations in pitch, tone, and volume to create interest and emphasize key points in my storytelling session.	B1
	I am able to control the speed of my narration to match the mood and action of the story.	B2
	I am able to speak clearly to ensure everyone can understand my story.	B3
	I am able to use facial expressions to convey emotions and reactions.	B4
	I am able to incorporate hand movements and body language to illustrate actions and events.	B5
	I am able to build connections with children during storytelling.	B6
	I am able to hold children's attention during storytelling.	B7
	I am able to involve the children in my storytelling session by asking questions, prompts, or incorporating activities.	B8
	I am able to use distinct voices for different characters to enhance the narrative.	B9
	I am able to highlight any underlying messages or lessons in the story.	B10
After storytelling	I am able to check children’s understanding of the story.	C1
	I am able to provide a satisfying ending that resolves the story effectively.	C2
	I am able to reinforce the story's lesson or theme.	C3
	I am able to encourage a discussion about the story to deepen understanding and engagement.	C4
	I am able to plan activities or crafts related to the story to extend the experience.	C5
	I am able to conduct activities or crafts related to the story to extend the experience.	C6

A pilot study was conducted involving 90 preschool teachers, selected using a random sampling technique, through an online questionnaire to assess the reliability and validity of the instrument. This number of respondents is sufficient to meet the recommended sample size requirements and achieve a 95% confidence level within the Rasch Measurement Model (Linacre, 1994). The Tunku Abdul Rahman University of Management and Technology (TAR UMT) Ethics Committee approved this study with reference number TAR UMT/EC/2024/02/20. The collected responses were analyzed using the Rasch Measurement Model through the WINSTEP software (Version 4.6.0) to evaluate the psychometric properties of the instrument. The procedures applied in this analysis included examining the reliability index and the respondent and item separation, identifying any polarization of construct-measuring items, and assessing item fit and misfit to ensure the accuracy and consistency of the measurement tool.

4. Results

The responses were analysed through three steps using the WINSTEP application to determine the reliability and separation index for item and respondent; item polarization through the value of Point-Measure Correlation (PTMEA-CORR) and check the item fitness through the MNSQ and ZSTD values of each item. The steps are explained in Figure 1 below.

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

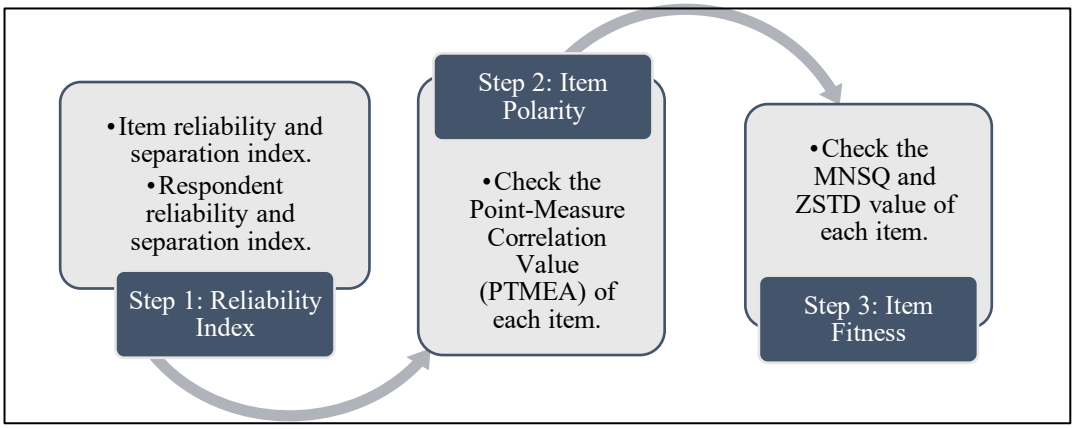

Figure 1. Steps for data analysis

4.1. Reliability Index

The Rasch Model required the person and item reliability and separation index to be considered in measurement instruments. Person or respondent reliability value is an assumption that the abilities of individuals in the sample are consistent despite being given a different set of items but still measuring the same construct. Meanwhile, item reliability is an indicator that the item is adequate for measuring the desired outcome (Toffoli et al., 2016). An index that divides respondents or items into multiple groups is known as a separation index.

According to Linacre (1994), an instrument is considered as reliable if the reliability index is 0.8 and above meanwhile the separation index is greater than 2.00. The result showed that the value of respondent reliability for the instrument is 0.94 and the separation index is 3.86. For item reliability, the result is 0.81 and the separation index is 2.08. Based on the result, this instrument has high respondent and item reliability and is able to divide respondents and items into multiple groups because it has a good separation index. The result is shown in the Table 5 below:

Table 5. Reliability and separation index for respondent and item

Person/Item	Reliability Index	Separation Index	N
Item	.81	2.03	24 items
Person	.94	3.86	90 persons

4.2. Item Polarity

One critical aspect of this Rasch model is item polarity, which refers to the direction of the relationship between item responses and the underlying trait being measured. Item polarity is analyzed through the Point Measure Correlation (PTMEA CORR) value in WINSTEP. Positive item polarity indicates that higher scores on an item correspond to higher levels of the trait, while negative polarity suggests the opposite. Positive polarity can be seen as a marker of good item quality. Items that align well with the underlying trait and show positive polarity are likely to be clear, relevant, and appropriately challenging for the target population. According to Boone and Noltemeyer (2017), if any value found in the PTMEA CORR section is a negative value, the items developed are not able to measure the construct that is intended to be measured. The item with the negative value needs to be checked or dropped from the

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

instrument. Table 6 shows the PTMEA CORR for all items in the instrument. Based on the result, all items showed positive value, and this indicates that all items are able to measure the competencies of teachers in conducting the storytelling.

Table 6. PTMEA CORR value for item polarity

Domain	Item Code	PTMEA CORR	Result
Before story telling	A1	0.69	Accepted
	A2	0.69	Accepted
	A3	0.82	Accepted
	A4	0.81	Accepted
	A5	0.78	Accepted
	A6	0.69	Accepted
	A7	0.74	Accepted
	A8	0.80	Accepted
During storytelling	B1	0.82	Accepted
	B2	0.80	Accepted
	B3	0.82	Accepted
	B4	0.85	Accepted
	B5	0.85	Accepted
	B6	0.85	Accepted
	B7	0.86	Accepted
	B8	0.84	Accepted
	B9	0.86	Accepted
	B10	0.82	Accepted
After storytelling	C1	0.83	Accepted
	C2	0.78	Accepted
	C3	0.82	Accepted
	C4	0.85	Accepted
	C5	0.78	Accepted
	C6	0.74	Accepted

4.3. Item Fitness Analysis

Item fitness refers to how well each item conforms to the expectations of the Rasch model. Item fitness is crucial for ensuring that the items accurately measure the underlying trait and contribute to a reliable and valid assessment. Item fitness is measured through the value of Infit and Outfit Mean Square (MNSQ). According to Bond and Fox (2013), for polytomous data, Infit and Outfit MNSQ values should be between 0.6 and 1.4 and for dichotomous data, the acceptable range is 0.7 to 1.3. Outfit MNSQ is more important than Infit MNSQ when determining how well items measure the construct. If the MNSQ value is above 1.4, the item is confusing and if the value is below 0.6, the item is too easy for respondents. Additionally, Infit and Outfit ZSTD values should be between -2.0 and 2.0. However, if the MNSQ values are within the acceptable range, the ZSTD values can be ignored.

Table 7 below shows the result of the item fitness of the instrument. The result shows that five items indicate Infit and Outfit MNSQ values out of the accepted range. The items outside the acceptable range are A1 (*I am able to choose appropriate story for preschool children*), A2 (*I am able to conduct my own*

https://doi.org/10.15405/epsbs.2026.09.77
Corresponding Author: Norazly Nordin
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

rehearsal before conducting storytelling to children), A5 (*I am able to produce the appropriate props for my storytelling session*), A6 (*I am able to understand the story before telling it to the children*) and A7 (*I am able to create vocal sound effects to create a more immersive experience in my story telling session*).

Table 7. Item fitness

Domain	Item Code	Infit MNSQ	Outfit MNSQ	Result
Before story telling	A1	1.64	1.59	Revised the item
	A2	1.74	1.86	Revised the item
	A3	0.97	0.93	Accepted
	A4	1.07	1.02	Accepted
	A5	1.51	1.41	Revised the item
	A6	1.18	1.47	Revised the item
	A7	1.45	2.17	Revised the item
	A8	1.19	1.08	Accepted
During storytelling	B1	0.83	0.84	Accepted
	B2	0.95	1.00	Accepted
	B3	0.81	0.71	Accepted
	B4	0.53	0.44	Accepted
	B5	0.60	0.59	Accepted
	B6	0.62	0.53	Accepted
	B7	0.69	0.63	Accepted
	B8	0.72	0.63	Accepted
	B9	0.75	0.86	Accepted
	B10	0.76	0.77	Accepted
After storytelling	C1	0.74	0.72	Accepted
	C2	1.01	1.04	Accepted
	C3	0.84	0.80	Accepted
	C4	0.64	0.63	Accepted
	C5	1.18	1.04	Accepted
	C6	1.36	1.21	Accepted

5. Conclusion

The process of assessing validity and reliability helps identify any weaknesses or flaws in the instrument. This allows for refinement and improvement, ensuring the instrument is as effective as possible. Validity ensures that the instrument measures what it is intended to measure. Without validation, the results obtained from the instrument may not accurately reflect the construct being studied. Meanwhile, reliability ensures that the instrument produces consistent and stable results over time. An unreliable instrument can lead to inconsistent findings, making it difficult to draw meaningful conclusions. This study employed the Rasch measurement model to evaluate the instrument's validity and reliability. The results demonstrated that the instrument is reliable, with all polarity items being positive. However, five items were identified as misfits. The impact of this study is significant as it provides a validated tool for assessing teacher competency in storytelling, which is crucial for enhancing educational practices. The identification of misfit items highlights areas for improvement, ensuring the instrument's precision and effectiveness. For future research, it is recommended to refine the misfit items and conduct further validation studies across diverse

educational settings to enhance the instrument's robustness and applicability. This will contribute to the continuous improvement of teacher competency assessments and support the development of effective storytelling techniques in education.

Acknowledgements

Special thanks to Tunku Abdul Rahman University of Management and Technology (TARUMT) for the award of the Internal Grant (Project No: UC/I/G2024-00132) to fund this research.

Author Contributions

Author A (Norazly Nordin) led the study's conceptualization and methodology development and took primary responsibility for writing the original draft. Author B (Joanne Lo Fui Senn) managed the data curation, performed the formal analysis using the Rasch Measurement Model, and created the visualizations. Author C (Muhammad Hilmi Abdul Rashid) was responsible for reviewing and editing the manuscript and validating the findings. All authors reviewed and approved the final manuscript.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Ethical Statement

Ethical approval for this study was obtained from Tunku Abdul Rahman University of Management and Technology (TAR UMT) Ethics Committee with reference number TAR UMT/EC/2024/02/20.

Funding

This study was supported by the Tunku Abdul Rahman University of Management and Technology (TAR UMT) under the TARUMT Internal Grant (Grant Number: UC/I/G2024-00132).

References

Archibald, D., Graham, C., & Larsen, R. (2020). Validating a blended teaching readiness instrument for primary/secondary preservice teachers. *Br. J. Educ. Technol., 52,* 536-551. https://doi.org/10.1111/bjet.13060

Arifin, M. A., Azizan, F. L., & Shahidan, A. N. (2021). A Needs Study Toward Developing Ma-TVETTCM Instrument to Measure TVET Teacher Competencies. *International Journal of Academic Research in Business and Social Sciences, 11*(6), 152-158. https://doi.org/10.6007/ijarbss/v11-i6/10104.

Aryadoust, V., Ng, L., & Sayama, H. (2020). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38,* 6-40. https://doi.org/10.1177/0265532220927487

Bidari, S. (2019). Story-based teaching: Activities for young learners. *Journal of Nelta, 24,* 233-236. https://doi.org/10.3126/nelta.v24i1-2.27695

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences.* Psychology Press.

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education, 4*(1), 1416898. https://doi.org/10.1080/2331186x.2017.1416898

Boone, W.J., Staver, J.R. (2020). Rasch Measurement Estimation Procedures. In: *Advances in Rasch Analyses in the Human Sciences* (pp. 187-198). Springer, Cham. https://doi.org/10.1007/978-3-030-43420-5_14

Burdelski, M., & Evaldsson, A. C. (2019). Young children's multimodal and collaborative tellings in family and preschool interaction. *Research on Children and Social Interaction, 3*(1-2), 1-5. https://doi.org/10.1558/rcsi.38982

Cooper, J. B., Lee, S., Jeter, E., & Bradley, C. L. (2020). Psychometric validation of a growth Mindset and Team Communication Tool to measure self-views of growth mindset and team communication skills. *Journal of the American Pharmacists Association, 60*(6), 818-826. https://doi.org/10.1016/j.japh.2020.04.012

Cruz, M., Saunders-Smits, G., & Groen, P. (2019). Evaluation of competency methods in engineering education: a systematic review. *European Journal of Engineering Education, 45,* 729 - 757. https://doi.org/10.1080/03043797.2019.1671810.

Eliyawati, E., Widodo, A., Kaniawati, I., & Fujii, H. (2023). The development and validation of an instrument for assessing science teacher competency to teach ESD. *Sustainability, 15*(4), 3276. https://doi.org/10.3390/su15043276

Filomeno, L., Forte, D., Di Simone, E., Di Muzio, M., Tartaglini, D., Lommi, M., & Ivziku, D. (2024). Systematic review and psychometric properties analysis of first-, middle-, and top-level nurse manager's core competencies instruments. *Journal of Nursing Management, 2024*(1), 2655382. https://doi.org/10.1155/2024/2655382.

Fischer, L., Rohm, T., Carstensen, C., & Gnambs, T. (2021). Linking of Rasch-Scaled Tests: Consequences of Limited Item Pools and Model Misfit. *Frontiers in Psychology, 12.* https://doi.org/10.3389/fpsyg.2021.633896.

Ghomi, M., & Redecker, C. (2019, May). Digital Competence of Educators (DigCompEdu): Development and Evaluation of a Self-assessment Instrument for Teachers' Digital Competence. In CSEDU (1) (pp. 541-548). https://doi.org/10.5220/0007679005410548

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18,* 66 - 67. https://doi.org/10.1136/eb-2015-102129.

Isbell, R., & Raines, S. (2013). *Creativity and the Arts with Young Children.* Wadsworth Cengage Learning.

Linacre, J. M (1994). Sample Size and Item Calibration Stability. [Artikel]. https://www.researchgate.net/publication/235361463_Sample_Size_and_ItemCalibration_Stability

Ludwikowska, K. (2022). Competency-Based Tests as a Tool for Teacher Evaluation in Higher Education Institutions. *Central European Management Journal, 30*, 85-111. https://doi.org/10.7206/cemj.2658-0845.83

Miller, S. D., & Pennycuff, L. (2008). The Power of Story: Using Storytelling to Improve Literacy Learning. *Journal of Cross-Disciplinary Perspectives in Education, 1*(1), 36-43.

Montenegro-Rueda, M., & Fernández-Batanero, J. (2023). Adaptation and validation of an instrument for assessing the digital competence of special education teachers. *European Journal of Special Needs Education*, *39*, 367 - 382. https://doi.org/10.1080/08856257.2023.2216573.

Nordin, N., Rossli, N. N., Ahmad Sabri, M. S., Nurdin, J., & Zulkuple, S. S. (2023). Validity and Reliability of the Survey Instrument of Preschool Teacher's Level of Knowledge Skills and Attitudes on 21st-Century Preschool Education. *Jurnal Penyelidikan Sains Sosial (JOSSR), 6*(20), 20 - 28.

Otoluwa, M., Talib, R., Tanaiyo, R., & Usman, H. (2022). Enhancing Children's VocabularyMastery Through Storytelling. JPUD - Jurnal Pendidikan Usia Dini. https://doi.org/10.21009/jpud.162.05.

Peláez-Sánchez, I., Glasserman-Morales, L., & Rocha-Feregrino, G. (2024). Exploring digital competencies in higher education: design and validation of instruments for the era of Industry 5.0. *Frontiers in Education*. https://doi.org/10.3389/feduc.2024.1415800.

Rusli, K. D. B., Chua, W. L., Ang, W. H. D., Ang, S. G. M., Lau, Y., & Liaw, S. Y. (2024). A hybrid systematic narrative review of instruments measuring home-based care nurses' competency. *Journal of Advanced Nursing, 80*(7), 2647-2671. https://doi.org/10.1111/jan.15904.

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2023). Interpreting results from Rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation, 46,* 591 - 603. https://doi.org/10.1080/09638288.2023.2169771

Toffoli, S. F. L., de Andrade, D. F., & Bornia, A. C. (2016). Evaluation of open items using the many-facet Rasch model. *Journal of Applied Statistics, 43*(2), 299-316. https://doi.org/10.1080/02664763.2015.1049938

Vetter, T., & Cubbin, C. (2019). Psychometrics: Trust, but Verify. *Anesthesia & Analgesia, 128,* 176–181. https://doi.org/10.1213/ANE.0000000000003859

Yaqoob Mohammed Al Jabri, F., Kvist, T., Azimirad, M., & Turunen, H. (2021). A systematic review of healthcare professionals' core competency instruments. *Nursing & health sciences, 23*(1), 87-102. https://doi.org/10.1111/nhs.12804

Wang, Y. (2024). The application effect of the Rasch measurement model combined with the CRF model: An analysis based on English discourse. *PLOS ONE, 19.* https://doi.org/10.1371/journal.pone.0309001