

Development and Rasch Validation of the Teacher Educators' Spiritual Values and Ethical Conduct Scale: A Pilot Study in Northwest Nigeria

Shetu Ilu Chetle¹, Adibah Abdul Latif²

Universiti Teknologi Malaysia, Faculty of Educational Sciences and Technology

DOI: <https://doi.org/10.47772/IJRISS.2026.100500808>

Received: 14 May 2026; Accepted: 19 May 2026; Published: 15 June 2026

ABSTRACT

Measurement in the social sciences and humanities often requires quantitative data to evaluate unobservable latent traits, particularly when assessing behavioral attributes. Ensuring validity and reliability is critical to maintain precision and accuracy in research instruments. This pilot study employed a quantitative survey design to evaluate the measurement properties of a self-assessment instrument designed to measure spiritual values and ethical conduct among teacher educators in Northwest Nigeria. The instrument, comprising 50 items, was administered to 37 teacher educators from Shehu Shagari College of Education. Content validity was assessed by six experts in psychology and education in Malaysia and Nigeria Universities and teacher training Institutions. Data were analyzed using Rasch Model Analysis (RMA) with Winsteps version 3.72.3, examining item reliability, item fit, unidimensionality, and item measurement properties. The findings indicated that content validity - Item-level scores (I-CVI) ranged from 0.72 to 0.80 across domains, while scale-level agreement (S-CVI/UA) reached 0.92 for spiritual values and 0.96 for ethical conduct. While several items demonstrated strong validity and reliability, others exhibited misfit statistics, suggesting the need for refinement or elimination prior to large-scale implementation. Specifically, items that failed to meet the criteria for acceptable fit and reliability, item targeting and local dependence were identified for modification leaving 41 items for the main thesis. This study underscores the potential of Rasch Model Analysis to enhance the psychometric robustness of self-assessment tools, ensuring cultural relevance and accuracy in measuring spiritual values and ethical conduct in educational contexts.

Keywords: Spiritual Values, Ethical Conduct, Self-assessment test, Rasch model analysis, Teacher Education.

INTRODUCTION

In an era marked by technological acceleration and individualism, the search for meaning and moral grounding has brought renewed scholarly attention to spirituality and ethics as foundational constructs (Gholam et al., 2023; Sylviana & Zahid, 2025). Spirituality is broadly conceived as the human capacity for connection to a transcendent reality, providing life with purpose, meaning, and internal tranquility beyond material concerns (Gholam et al., 2023; Kumar et al., 2017). Closely intertwined, ethical behaviour constitutes the tangible expression of this inner framework, manifested through actions aligned with societal virtues such as integrity, fairness, and dignity (Kumar et al., 2017). This conceptual pairing posits spirituality as the internal, value-generating dimension and ethics as its external, socially observable correlate—a relationship with profound implications for holistic education, which aims to develop the whole person within a social context (Uknowledge & Bower, n.d.). Teacher spirituality in this study reflects values such as tolerance, trust, altruism, respect, and honesty, and extends beyond religion to encompass meaning-making, transcendence, and a sense of connectedness (McGhee & Grant, 2008; Pong, 2022). The ethical conduct in the context of the Nigeria's code of conduct as featured in this study are: obligation, integrity, law-abiding behaviour, role modelling and compassion (FRN, 2013).

The significance of spiritual values and ethical conduct in teacher education has garnered increased attention in recent years, as these attributes play a pivotal role in shaping the moral and professional character of future

educators (Aydin, 2020; Kumar, 2022). Teacher educators are responsible not only for imparting academic knowledge but also for modeling ethical behavior and instilling values that promote integrity and responsibility among students (Vem, 2020). The integration of spiritual values into teacher education has been linked to improved teacher well-being, job satisfaction, and positive student outcomes (Korthagen & Vasalos, 2005; Tal & Morag, 2009).

Despite the growing recognition of these values, there remains a critical need for reliable and valid assessment instruments that can measure spiritual values and ethical conduct among teacher educators. Existing research has often focused on student populations or has been limited to Western and Asian contexts, thereby neglecting cultural considerations pertinent to Nigerian teacher educators (Salleh, 2017; Kisiel, 2018). Furthermore, while self-assessment tests are frequently used in teacher education to evaluate affective traits, concerns about their psychometric robustness persist. Specifically, studies have highlighted gaps in the rigorous validation of these instruments using modern psychometric techniques, such as the Rasch Model Analysis (Talip, 2019; Rosli et al., 2020).

The Rasch Model, a probabilistic model of item response theory, offers a powerful method for evaluating the psychometric properties of assessment tools, transforming ordinal data from Likert scales into interval-level data that provide more precise and meaningful measurements (Boone & Staver, 2014; Bond, 2013). This approach is crucial in ensuring that instruments are both culturally responsive and psychometrically sound, especially in diverse educational settings like Nigeria, where cultural context significantly influences values and ethical conduct (Morag, 2020; Tal, 2019).

This study aims to validate a self-assessment instrument for measuring spiritual values and ethical conduct among teacher educators in Northwest Nigeria. By employing Rasch Model Analysis, the research seeks to address the pressing need for culturally relevant and psychometrically robust assessment tools in the field of teacher education. This research will be answering research questions like: What is the extent of the instrument validity? How reliable are the items? Does the instrument meet Item-person targeting? The findings will contribute to the development of more reliable and valid instruments, ultimately enhancing the assessment and cultivation of ethical and spiritual values among educators.

METHODOLOGY

In this study, the researchers employed a cross-sectional survey design, to explore the spiritual values and ethical conduct of teacher educators. Data were collected at a single point in time, offering a practical and efficient means of capturing a broad snapshot of educators' perceptions, behaviors, and instructional practices. To ensure the robustness of the instrument, a pilot study was conducted. This preliminary phase involved testing item reliability and assessing how well each item aligned with the Rasch measurement model. Additionally, participant feedback was carefully integrated to improve both the linguistic clarity and cultural relevance of the scale items, making the tool more contextually appropriate for the target population.

The scale questionnaire is structured into two main domains: Spiritual Values and Ethical Conduct. Section "A" comprises twenty-five constructs related to Spiritual Values, while Section "B" includes twenty-five constructs focused on Ethical Conduct. Each construct is assessed through five items, resulting in a total of 50 items formatted on a four-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree). Notably, negatively worded items were reverse-scored to maintain consistency. This scale format is widely accepted in social science research for evaluating latent traits (Hayat et al., 2020).

To validate the content of the Self-Assessment scale questionnaire, a rigorous content validation process was conducted. Six experts specializing in psychology and educational assessment from universities in Malaysia and Nigeria were selected based on their extensive experience and academic qualifications in the relevant fields. These experts evaluated the content of the 50 items to ensure clarity, relevance, and cultural appropriateness (Hair et al., 2018). The level of relevance was determined using Content Validity index such as Item content validity index I-CVI and scale content validity index S-CVI based on Lynn 1986 and Polit and Beck, 2006 thresholds.

Following the content validation, the questionnaire was distributed to 41 teacher educators from Shehu Shagari College of Education, located in Northwest Nigeria. This college was chosen as a representative site for the study, although it was not included in the final sample to avoid bias. A total of 37 completed questionnaires were retrieved and subsequently analyzed using Winsteps software version 3.72.3. The analysis employed Rasch Model Analysis (RMA) to assess item reliability, item fit, unidimensionality, and overall measurement quality while Cronbach Alfa was applied for internal consistency.

DATA ANALYSIS / FINDINGS

Content validity

In the pilot study, Content validity index (CVI) was used for analysing the expert agreement. Some items were reviewed or reworded without any dropping. Score on each item was provided and CVI was obtained by calculating both I-CVI and S-CVI.

Item Reliability / Person Reliability:

The pilot study result showed a Cronbach Alpha (α) reliability of 0.87 indicating its internal consistency reliability. Pilot study analysis revealed that, the item reliability has value of 78, which indicates the items are moderate in terms of size and difficulty to assess the teacher educators' spiritual values and ethical conduct. The person reliability index is strong (0.84), which implies that the instrument is sufficiently able to differentiate between high and low spiritual values and ethical conduct educators.

Item Polarity

The Point Measure Correlation (PTMEA CORR), quantifies the degree to which each item functions consistently with the overall trait being measured. This thereby providing empirical evidence for construct validity (Müller, 2020; Rosli et al., 2020). As noted by Boone et al. (2014), PTMEA CORR values serve as a primary diagnostic for item fit within Rasch-based measurement frameworks. The interpretive benchmarks commonly cited in the psychometric literature suggest that PTMEA CORR values exceeding 0.40 indicate exceptionally strong item–construct congruence, values between 0.30 and 0.39 denote good alignment, and those ranging from 0.20 to 0.29 reflect fair performance.

Conversely, values falling between 0 and 0.19 signal a failure of the item to adequately discriminate among respondents, undermining its utility. Moreover, for an item to retain polarity, its PTMEA CORR must be positive, demonstrating that higher scores on the item are systematically associated with higher aggregate scores on the remaining items (Allen et al., 2022). Applying these criteria to the instrument, the following items were identified as exhibiting positive yet suboptimal discrimination: 15, 3, 50, 33, 8, 7, 20, 5, 4, and 47. Nevertheless, it is important to note that the vast majority of TEDSAT items demonstrated PTMEA CORR values in the extremely good range, indicating robust construct alignment overall.

Item-measure:

Item-measure analysis employs fit statistics—specifically infit and outfit Mean-Square (MNSQ) values—to rank items hierarchically and ensure alignment with respondent ability levels when measuring constructs such as spiritual values and ethical conduct. According to Bond and Fox (2007), acceptable MNSQ values range from 0.6 to 1.4; values exceeding 1.4 indicate misfit, while those below 0.6 suggest random guessing by respondents. Standardized Z-scores (ZSTD) serve as a supplementary diagnostic with acceptable bounds between -2 and $+2$ (Bond & Fox, 2015), though ZSTD may be disregarded when MNSQ is acceptable (Linacre, 2018). Pilot study findings revealed several misfitting items: CO1 (MNSQ = 1.50), RT5 (MNSQ = 1.50, ZSTD = 2.4), and RS5 (MNSQ = 1.66, ZSTD = 2.8), all exceeding established thresholds. Consequently, systematic refinement of these items is warranted to enhance construct validity and measurement precision.

Item Unidimensionality

In Rasch-based measurement, the unidimensionality of an instrument is empirically evaluated using Residual Principal Component Analysis (PCA), which serves as a critical diagnostic for ensuring that a single latent trait underlies responses to the items (Boone et al., 2014; Rosli et al., 2020). Unidimensionality is not merely a statistical nicety but a fundamental prerequisite: if this condition is violated, aggregating total scores or making meaningful comparisons between respondents or items becomes statistically indefensible (Boone et al., 2014). To establish unidimensionality, the variance explained by the measures must be at least 20% (Bond & Fox, 2007). Furthermore, the unexplained variance observed in the first contrast which represents the largest residual dimension after extracting the primary construct should not exceed 15%, as exceeding this threshold suggests the presence of a secondary dimension that threatens construct validity (Rosli et al., 2020).

A unidimensional construct measures what it claims to measure (no presence of other variance outside of the construct). The unidimensional index measures the instrument based on Residual Principal Component Analysis (PCA). According to Linacre (2018 as cited in Siew et al., 2019; Rosli et al., 2020), the minimum variance to show the unidimensionality of the instrument is 20%.

Standardized Residual variance (in Eigenvalue units)

	-- Empirical --		Modeled
Total raw variance in observations =	65.6	100.0%	100.0%
Raw variance explained by measures =	15.6	23.8%	25.3%
Raw variance explained by persons =	4.4	6.6%	7.1%
Raw Variance explained by items =	11.2	17.1%	18.2%
Raw unexplained variance (total) =	50.0	76.2%	74.7%
Unexplnd variance in 1st contrast =	7.4	11.4%	14.9%
Unexplnd variance in 2nd contrast =	5.1	7.7%	10.1%
Unexplnd variance in 3rd contrast =	4.2	6.5%	8.5%
Unexplnd variance in 4th contrast =	3.1	4.7%	6.2%
Unexplnd variance in 5th contrast =	2.9	4.4%	5.8%

The result above indicates that the raw variance explained by the measures in this pilot study is 23.8% and it's 3.8% above the expected minimum percentage of 20%. The unexplained variance in the first contrast should not exceed 15% (Fisher, 2007 as cited in Rosli et al., 2020). The pilot study result gives 11.4% as the unexplained variance in the first, second was 7.7%, third was 6.5%, fourth was 4.7%, and fifth contrasts was 4.4%; all of which did not exceed 15% nor any of the lower contrasts exceeding the previous. Based on Rating Scale Instrument Quality Criteria Fisher in table 2 above, the pilot test is said to be good, as the 2nd to 5th fall between (5-10%). The pilot test instrument by the standard of unidimensionality is relevant for measuring the Spiritual Values and Ethical Conduct of teacher educators in Northwest, Nigeria.

Item Wright Map

Item Wright Map or the variable map or Wright map from Rasch measurement, has the distinct benefit of aligning person performances as well as item performances on the same interval scale (Boone, 2016). The performance on the item-person map as on the vertical axis, the figure locates items according to item difficulty, while the size of the item marker indicates the relative precision (standard error) of that estimate (Wright & Mok,

2004). On the horizontal axis, the fit of those items is estimated in a standardized form as a t distribution and easier items are located at the bottom of the pathway, with more difficult items at the top (Boone, 2016). Items which misfit the model are located to the right; items which are closer to the deterministic.

RMA identifies the correlation between teacher educators' spiritual values and ethical conduct and self-assessment item difficulty using a logarithmic function. Teacher educators' latent traits in spiritual values and ethical conduct are highly related to the educators' moral stand and item difficulty (Boone, et al, 2014; Wright & Mok, 2004). As revealed in the Person-Item Map which indicates teacher educators' spiritual values and ethical conduct and item difficulty in the logit scale ranging from -1 to 2 . The top left shows educators who have high spiritual values and ethical conduct, while the bottom left indicates educators with lower spiritual values and ethical conduct. Item difficulty is shown on the right side of the map. Teacher educators hardly answer "agree" to CO1 CO2 CO3 TR1, while they easily answer "agree" to AL5.

Major Findings

Content Validity

The assessment tool showed strong content validity. Item-level scores (I-CVI) ranged from 0.72 to 0.80 across domains, while scale-level agreement (S-CVI/UA) reached 0.92 for spiritual values and 0.96 for ethical conduct—all exceeding recommended minimums. This indicates a solid expert consensus on the relevance and clarity of the items, confirming the Scale as a psychometrically sound instrument aligned with established validation standards (Connell et al., 2018; Roebianto et al., 2023).

The results of the Rasch Model Analysis provided comprehensive insights into the psychometric properties of the TEDSAT questionnaire. The analysis focused on item reliability, person reliability, item fit, unidimensionality, and the Item Wright Map to evaluate the instrument's effectiveness in measuring spiritual values and ethical conduct among teacher educators.

Item Reliability and Item Person Reliability

The item reliability was calculated at 0.78 , indicating moderate consistency, while the person reliability score was 0.84 , suggesting a strong ability of the instrument to distinguish between respondents with varying levels of spiritual values and ethical conduct. The Cronbach's Alpha for internal consistency was 0.87 , which exceeds the acceptable threshold of 0.70 , further confirming the instrument's reliability. The person and item separation indices were 2.29 and 1.90 , respectively, indicating an acceptable level of distinction between high and low performers.

Table 1. Item Reliability and Person Reliability

Reliability Measure	Value	Interpretation
Item Reliability	0.78	Moderate
Person Reliability	0.84	Strong
Cronbach's Alpha	0.873	High Consistency
Item Separation Index	1.90	Moderate
Person Separation Index	2.29	Strong

Item Polarity/ Item Fit

Point Measure Correlation (PTMEA CORR) values were assessed to determine item polarity, with some items showing acceptable correlations above 0.20 while lower than 0.20 and negative PTMEA CORR were to be refined such as items 15, 3, 50, 33, 8, 7,20, 5, 4 and 47. Item fit statistics were evaluated using Mean Square (MNSQ) infit and outfit values, which should ideally range between 0.6 and 1.4 (Bond & Fox, 2007). Items such as 15, 3, 50, 33, 8, and 7 demonstrated misfit, suggesting the need for refinement or possible elimination.

Unidimensionality

The Principal Component Analysis (PCA) of residuals indicated that 23.8% of the variance was explained by the measures, which exceeds the minimum requirement of 20%, supporting the unidimensionality of the instrument. The unexplained variance in the first contrast was 11.4%, well below the 15% threshold, indicating that the instrument measures a single construct effectively.

Table 3. Unidimensionality Analysis

Variance Measure	Percentage	Threshold	Interpretation
Variance Explained by Measures	23.8%	>20%	Unidimensionality Met
Unexplained Variance (1st Contrast)	11.4%	<15%	Acceptable Unidimensionality

Item Wright Map

The Item Wright Map, also referred to as a person-item map in Rasch terminology offers a graphical depiction of how item difficulty calibrations and respondent ability estimates are situated conjointly along the same logit scale, thereby enabling an intuitive assessment of measurement targeting. In the present analysis, the map revealed a broad distribution of teacher educators' abilities, suggesting substantial heterogeneity in the sample with respect to the underlying constructs of spiritual values and ethical conduct. Most items demonstrated appropriate coverage across the continuum, indicating satisfactory alignment between item difficulty and person ability for the majority of respondents. Nonetheless, notable gaps were observed, particularly at the upper region of the scale, where few or no items were calibrated to match the highest levels of spiritual and ethical functioning.

This empirical observation points to a ceiling effect in the current instrument, underscoring the need for additional items of greater difficulty to more precisely discriminate among and accurately assess respondents who manifest exceptionally high spiritual and ethical values.

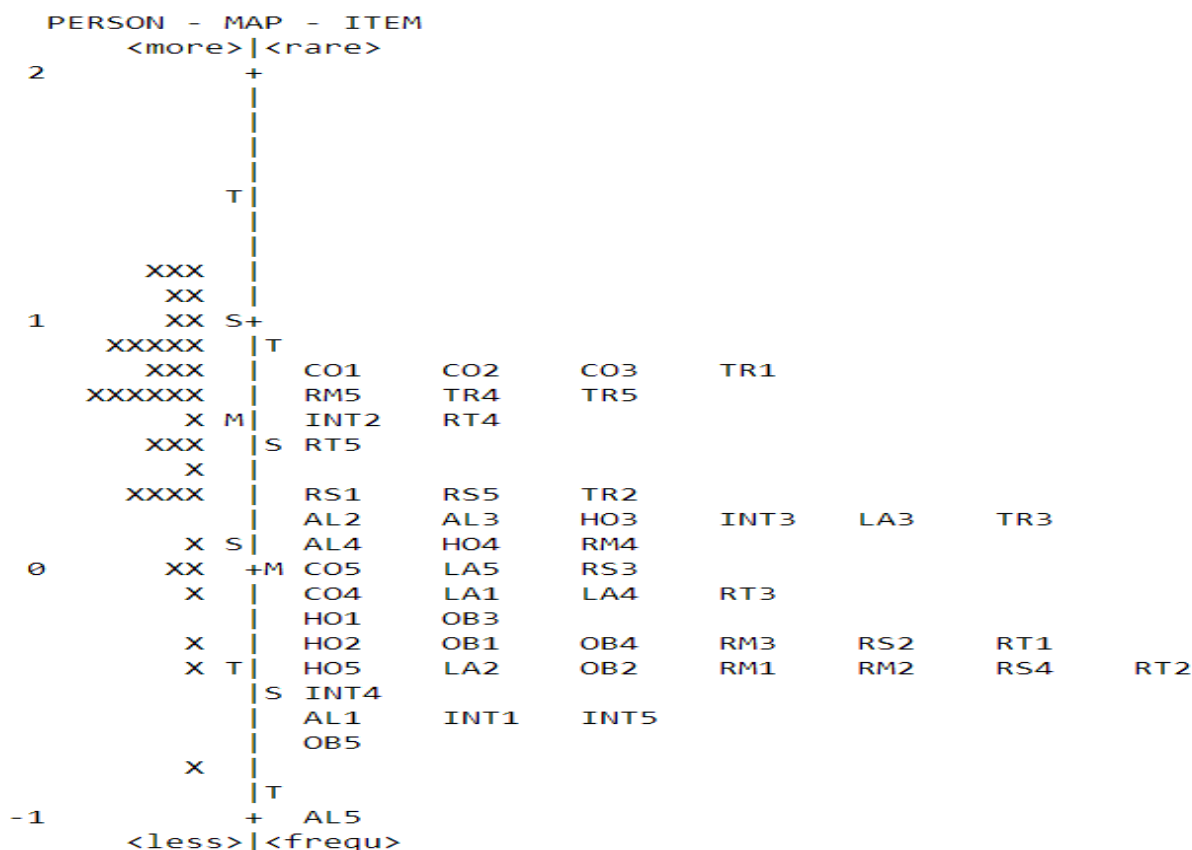


Figure 1. Person-Item Map

The Rasch Model is a probabilistic approach used to measure latent traits such as ability or perception by analyzing person-item interactions (Bond & Fox, 2019). The "Person-Item Map" framework provided aligns closely with Rasch Model principles by visually representing participants' ability levels and item difficulty on a common scale. Rare items such as "CO1," "CO2," and "CO3" occupy the topmost level items are suitable for differentiating high-performing participants (Linacre, 2020; Bond et al., 2021). Items in this range often provide the most information about the population under study (Boone et al., 2020). Future studies could expand on this by exploring multidimensional Rasch models to capture complex traits as the map indicate deficiency in high profile items at the top for highly spiritual and ethical teacher educators

The pilot study confirmed that the Scale is a sound instrument, showing good reliability, acceptable item performance, unidimensional structure, and overall construct validity. A few items with poor fit or weak discrimination (RT5, TR3, RS5, AL5, HO4, TR1, CO1, CO2, and OB1) were flagged off leaving 41 items. Addressing these issues will enhance the scale's precision and readiness for the main study.

DISCUSSION

The global perspective highlights several best practices in self-assessment that contribute to its effectiveness in educational systems and provides clear assessment criteria and expectations, offering timely and constructive feedback, promoting self-reflection and goal-setting, and fostering a supportive learning environment (Arar & Saiti, 2022). When implemented effectively, self-assessment enhances student motivation, engagement, and academic achievement. This is however, a challenge in Nigerian educational system where it is lamented that "It is a well-known fact that there is inadequate knowledge of the theory and practice of educational assessment techniques among lecturers in higher institutions in Nigeria, which is rather unfortunate particularly, Psychometrics which deals with design and interpretation of tests that measure psychological traits such as values, attitudes, ethics, personality, motivation, happiness, and intelligence" FRN, 2022 pp. iii. This calls for competence in value assessments, perform high levels of teaching skills, meet accountability standards, share professional knowledge, care about students and their success, and hold distinctive qualities that characterize their effectiveness in developing instruments for assessment and validation in the affective domain (George, 2022; Salleh, 2017; Setiawan et al., 2019).

The content validity indices obtained in this pilot study (I-CVI: 0.72–0.80; S-CVI/UA: 0.92–0.96) and Cronbach's alpha reliability (0.873) closely align with the findings of Chetle and Latif (2025), who reported I-CVI values ranging from 0.72 to 0.92 and a Cronbach's alpha of 0.70 during their psychometric validation of the TEDSAT Scale among Nigerian teacher educators. The person separation index of 2.29 and person reliability of 0.84 obtained in this study meet the acceptable criteria established by Fisher (2007), who recommends person separation values of at least 2.0 and person reliability above 0.8 for adequate instrument performance, while the item reliability of 0.78 falls within the "good" range of 0.81–0.90 as per Boone, Staver, and Yale (2014).

Regarding unidimensionality, the Principal Component Analysis of residuals revealed that 23.8% of the variance was explained by the measures, exceeding the minimum 20% threshold, while the unexplained variance in the first contrast was 11.4%, below the 15% upper limit recommended by Linacre (2012) and Fisher (2007) to indicate that no meaningful secondary dimension exists. The identification of some misfitting items and items with weak polarity (near-zero PTMEA CORR values) aligns with the findings of Bond and Fox (2015), who note that such items are commonly flagged during pilot testing and require refinement or removal before main study implementation to ensure optimal scale functioning.

This finding is significant as it confirms the theoretical basis that spiritual values and ethical conduct can be measured as a unified construct, aligning with prior studies (Korthagen & Vasalos, 2005; Tirri, 2019).

The Item Wright Map provided additional evidence of the instrument's effectiveness, displaying a well-distributed range of item difficulty levels. However, the identified gaps at the higher end of the scale suggest that future versions of the TEDSAT questionnaire may benefit from adding more items to better capture the upper spectrum of spiritual and ethical values among teacher educators. This aligns with previous researches that highlights the importance of comprehensive coverage in psychometric instruments (Boone, 2016; Yasin et al., 2015).

Implications For Practice and Policy

The results of this study have significant implications for the field of teacher education. By providing a validated and reliable tool for assessing spiritual values and ethical conduct, educational institutions can more effectively evaluate and promote these critical attributes in teacher preparation programs. Policymakers can also use the instrument to inform the development of teacher training curricula that emphasize ethical and spiritual growth.

In essence, the scale indicated strong validity and reliability evidence, but it is not yet optimally calibrated. The implications point toward targeted refinement of the misfit items, with careful interpretation of individual scores, and confidence in its use for group-level assessment of spiritual and ethical values within the Nigerian Colleges of Education.

LIMITATION OF THE STUDY

The sample size was really small, testing a 50-item survey on just 37 people is fine for a pilot, but it's not nearly enough to draw firm conclusions. With such a small group, the results might not hold up if the same survey is conducted with a larger population. Moreover, the study only focused on Northwest which limits generalization of the results. Although, it can provide insight to researchers investigating other parts of Nigeria. There is possibility of socio-cultural desirability bias in the responses even though Rasch Measurement Model is robust statistical tool to detect and manage such.

CONCLUSION

In conclusion, 41 out of 50 items were considered for the main thesis. This provides a strong foundation for ongoing efforts to develop reliable, culturally responsive self-assessment tool in affective domain of education. The integration of rigorous psychometric techniques, such as Content Validity Index, Cronbach Alpha value and Rasch Measurement analysis, in instrument validation represents a significant advancement in the field, ensuring that measures of spiritual and ethical values are both accurate and meaningful across diverse educational contexts. What truly sets this study apart is that it didn't just test a scale—it brought Rasch measurement theory into a real-world, culturally grounded setting. Most psychometric research tends to overlook non-Western or context-specific educational environments, leaving a noticeable gap in how we validate self-assessment tools for teacher educators in places like Northwest Nigeria. This study steps right into that gap

RECOMMENDATIONS FOR PRACTICE

The validated TEDSAT instrument offers practical benefits for educational institutions and policymakers. It can be employed to assess and enhance teacher training programs, ensuring that spiritual values and ethical conduct are given due attention in the curriculum. Institutions may use this tool to identify areas where educators require further development, ultimately fostering a more ethically aware teaching workforce.

Future Research Directions

Despite the study's promising results, future research should aim to test the instrument with larger and more diverse samples to improve generalizability. Cross-cultural validation studies would also strengthen the global applicability of the TEDSAT questionnaire. Additionally, further refinement of misfit items and the inclusion of more items to cover higher levels of spiritual and ethical values are recommended to enhance the instrument's comprehensiveness and precision. Future researchers can replicate the study by employing Exploratory Factor Analysis and Confirmatory Factor Analysis to see if the same results can be obtained. Mixed method can be adopted to deep into the relationship between the spiritual values and ethical conduct. Similarly, longitudinal survey can be employed to determine if the relationship can be sustained. This study did not observe demographic factor of the respondents; other researchers can investigate into the demographic factor for Differential Item Functioning.

REFERENCES

1. Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single Item Measures in Psychological Science: A Call to Action. In *European Journal of Psychological Assessment* (Vol. 38, Issue 1, pp. 1–5). Hogrefe Publishing GmbH. <https://doi.org/10.1027/1015-5759/a000699>
2. Aydin, A. (2020). Spiritual values in teacher education. *Journal of Teacher Education*, 71(1), 5-18.
3. Arar, K., and Saiti, A. (2022). Ethical leadership, ethical dilemmas and decision making among school administrators. *Equity in Education and Society*, 1(1), 126–141. <https://doi.org/10.1177/27526461211070828>
4. Boone, W. J. (2016). Rasch analysis for instrument development: Why, when and how? *CBE Life Sciences Education*, 15(4). <https://doi.org/10.1187/cbe.16-04-0148>
5. Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Person Reliability, Item Reliability, and More. In *Rasch Analysis in the Human Sciences* (pp. 217–234). Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4_10.
6. Boone, W. J., Staver, J. R., & Yale, M. S. (2020). *Rasch Analysis in the Human Sciences*. Springer
7. Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* Second Edition. www.erlbaum.coOl
8. Bond, T. G. (2013). Applying the Rasch model: Fundamental measurement in the human sciences. *Journal of Applied Measurement*, 14(1), 1-15.
9. Bond, T. G., & Fox, C. M. (2019). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (4th ed.). Routledge
10. Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model in Social Sciences Research*. RoutledgeCampbell, E. (2020). Ethical conduct in teacher education. *Journal of Educational Ethics*, 19(1), 1-12.
11. Linacre, J. M. (2020). A User's Guide to WINSTEPS Rasch-Model Computer Programs. Winsteps.com
12. Faradillah, A., Jusra, H., & Pradipta, T. R. (2021). Validity and Reliability Analysis of Algebraic Reasoning Test Instrument. *Journal of Physics: Conference Series*, 1933(1). <https://doi.org/10.1088/1742-6596/1933/1/012116>
13. Fox, C. M. (2017). Applying the Rasch model: Fundamental measurement in the human sciences. *Journal of Applied Measurement*, 18(2), 1-15.
14. FRN. (2022). *Manual of Critical Assessment Competences for Tertiary Education Lecturers In Nigeria*.
15. FRN. (2013). *CODE OF CONDUCT for Teaching Profession in Nigeria*
16. George, J. A. (2022). Assessing Affective Domain In Higher Education: Educator's Perception. <https://doi.org/10.17605/OSF.IO/ASD49>
17. Gholam, G., Poorkami, H., Fallah, V., & Salimi, L. (2023). The Role of Ethical Values in the Meaning of Life and Its Impact on Education from the View of Frankl's Approach. *Int J Ethics Soc*, 5(2), 27–34. <https://doi.org/10.22034/ijethics.5.2.41>
18. Kisiel, J. (2018). Self-assessment tests in teacher education. *Journal of Teacher Education*, 69(2), 139-153
19. Korthagen, F. A. J. (2017). Teacher educator professionalism. *Journal of Education for Teaching*, 43(1), 5-18
20. Krishnan, S., & Idris, N. (2014). Investigating Reliability and Validity for the Construct of Inferential Statistics. In *International Journal of Learning, Teaching and Educational Research* (Vol. 4, Issue 1).
21. Kumar, S. (2022). Spiritual values development in teacher education. *Journal of Spiritual Education*, 15(1), 1-15.
22. Kumar, S., Basava, G., & Professor, A. S. (2017). "The Influence of Spiritual Intelligence on Ethical Behaviour in ITES Organizations with respect to Bengaluru." In *International Journal of Research and Scientific Innovation (IJRSI) |: Vol. IV*. www.rsisinternational.org
23. Linacre, J. M. . (2007). *A user's guide to WINDTEPS Rasch-model computer programs*. Chicago, Illinois: MESA Press.
24. Morag, O. (2020). Cultural context and spiritual values in teacher education. *Journal of Multicultural Education*, 44(1), 1-15.
25. Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, 7(1). <https://doi.org/10.1186/s40488-020-00108-7>

26. Rosli, R., Abdullah, M., Siregar, N. C., Abdul Hamid, N. S., Abdullah, S., Beng, G. K., Halim, L., Daud, N. M., Bahari, S. A., Majid, R. A., and Bais, B. (2020). Student Awareness of Space Science: Rasch Model Analysis for Validity and Reliability. *World Journal of Education*, 10(3), 170. <https://doi.org/10.5430/wje.v10n3p170>
27. Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, 14(3). <https://doi.org/10.29333/iejme/5755>
28. Salleh, A. (2017). Measuring Values in Modern School System. *International Journal of Modern Education Studies*, 1(1), 28. <https://doi.org/10.51383/ijonmes.2017.11>
29. Siew, N. M., Syafiq, M., & Rahman, A. (2019). Assessing the Validity and Reliability of the Future Thinking Test using Rasch Measurement Model. In *International Journal Of Environmental & Science Education* (Vol. 14, Issue 4). <http://www.ijese.com>
30. Setiawan, A., Mardapi, D., Supriyoko, and Andrian, D. (2019). The development of instrument for assessing students' affective domain using self- and peer-assessment models. *International Journal of Instruction*, 12(3), 425–438. <https://doi.org/10.29333/iji.2019.12326>
31. Sylviana, E., & Zahid, B. M. (2025). Strengthening Spiritual Development Approaches in Teaching and Learning for Malaysian Higher Education Students: An Integrative Conceptual Framework. *INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE(IJRISS)*, ix(VI), 3819–3827. <https://doi.org/10.47772/IJRISS>
32. Tal, T. (2019). Cultural context and teacher education. *Journal of Teacher Education*, 70(1), 5-18.
33. Talib, R., Iahad, N. A., Ashari, Z. M., Rameli, M. R. M., Bakar, Z. A., & Dollah, R. (2019). Rasch strategies for evaluating quality of the Conceptions and Alternative assessment Survey (CETAS). *Universal Journal of Educational Research*, 7(12 A), 10–17. <https://doi.org/10.13189/ujer.2019.071902>
34. Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1. The “most likely” measures coming from the model. *Disability and Rehabilitation*, 46(3), 591–603. <https://doi.org/10.1080/09638288.2023.2169771>
35. Tirri, K. (2019). Ethical conduct development in teacher education. *Journal of Educational Ethics*, 18(1), 1-12.
36. Uknowledge, U., & Bower, W. C. (n.d.). Moral and Spiritual Values in Education: A Challenge to Every Moral and Spiritual Values in Education: A Challenge to Every American American. https://uknowledge.uky.edu/upk_social_and_philosophical_foundations_of_education/2
37. Vem, L. J., Tuamyil, R. B., Ocholi, S. A., and Thurasamy, R. A. (2020). Turnover Intention in Nigerian Universities: Do Academics' Spirituality and the Spiritual Climate Matter? *Africa Education Review*, 17(6), 65–85. <https://doi.org/10.1080/18146627.2022.2063750>
38. Wright, B. D., & Mok, M. M. C. (2004). An Overview of the Family of Rasch Measurement Models.
39. Wright, B. D., & Stone, M. H. (2018). *Best Test Design: Rasch Measurement*. Mesa Press
40. Yasin, R. Mohd., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204, 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>