Diversitas Journal
ISSN 2525-5215
Volume 11, Issue 1 (Jan./Mar. 2026) p. 42 – 57
https://diversitasjournal.com.br/diversitas_journal

Gender-Based Analysis of Teacher Empowerment Scale: Examining Factor Structure and Rasch Model Fit in Higher Education

PANELA, Teody Lester V. [1]

[1] 0000-0002-2251-8949; Northwest Samar State University, Philippines. teodylester.panela@nwssu.edu.ph

A B S T R A C T

This study examined the measurement invariance and psychometric properties of the Teacher Empowerment Scale across gender groups in higher education. Using Rasch analysis, 86 items spanning three factors (fostering continuous improvement, teaching ownership and freedom, and work climate and conditions) were analyzed with data from 968 faculty members. Results demonstrated excellent model fit (mean infit/outfit MNSQ ≈1.00) and high reliability (α=0.90-0.93) across all factors. Differential item functioning analysis revealed minimal gender-based variations, with only 5 items in factor 1, 4 items in factor 2, and none in factor 3 showing significant differences. The scale provides fair assessment of teacher empowerment constructs for both male and female educators, supporting previous research findings. Recommendations include implementing the scale confidently while attending to items with differential functioning; refining these items to enhance gender neutrality; extending validation research to additional demographic variables; conducting longitudinal studies; and utilizing the three-factor structure for designing targeted interventions. This research addresses existing gaps regarding gender considerations in scale development, advancing equitable assessment instruments for higher education settings.

R E S U M O

Este estudo examinou a invariância de medição e as propriedades psicométricas da Escala de Empoderamento de Professores entre grupos de gênero no ensino superior. Usando a análise Rasch, 86 itens abrangendo três fatores (promoção da melhoria contínua, propriedade e liberdade docente e clima e condições de trabalho) foram analisados com dados de 968 membros do corpo docente. Os resultados demonstraram excelente ajuste do modelo (média infit/outfit MNSQ ≈1,00) e alta confiabilidade (α=0,90-0,93) em todos os fatores. A análise diferencial do funcionamento dos itens revelou variações mínimas baseadas no género, com apenas 5 itens no fator 1, 4 itens no fator 2 e nenhum no fator 3 apresentando diferenças significativas. A escala fornece uma avaliação justa dos construtos de capacitação de professores tanto para educadores masculinos como femininos, apoiando resultados de pesquisas anteriores. As recomendações incluem implementar a escala com confiança ao mesmo tempo em que atende itens com funcionamento diferencial; aperfeiçoar estes itens para aumentar a neutralidade de género; estender a pesquisa de validação a variáveis demográficas adicionais; realização de estudos longitudinais; e utilizar a estrutura de três factores para conceber intervenções específicas. Esta investigação aborda as lacunas existentes relativamente às considerações de género no desenvolvimento da escala, promovendo instrumentos de avaliação equitativos para ambientes de ensino superior.

ARTICLE INFORMATION

Article process:
Submitted: 03/15/2025
Approved: 01/06/2026
Published: 01/22/2026

Keywords:
Measurement invariance, Item response theory, Construct validity, Faculty development, Organizational climate

DOI: 10.48017/dj.v11i1.3370

Introduction

Accurate measurement of teacher empowerment across demographic groups remains fundamental to equitable educational assessment, particularly as institutions navigate post-pandemic educational transformations. While teacher empowerment's significance in educational settings is well-established, gender influences on measurement tools require careful examination to ensure valid comparisons across diverse faculty populations.

Contemporary research indicates gender's substantial role in how teachers experience and express empowerment (Balkar, 2015; Chebet, 2014). However, limited research examines measurement invariance of teacher empowerment scales across gender groups, particularly in higher education contexts where organizational structures and professional expectations may interact differently with gender dynamics. This gap raises critical questions about measurement tool validity when applied across different gender groups.

Teacher empowerment encompasses multidimensional constructs involving various aspects of educators' professional lives. Alvunger and Wahlström (2018) characterize empowered teachers as professionals capable of exercising autonomous judgment, effectively managing challenges, and adapting to institutional changes. The construct encompasses leadership and decision-making, professional development, reputation and standing, teacher efficacy, institutional autonomy, collegiality, and work climate dimensions (Simon, 2013; Oracion, 2015).

Research demonstrates that teacher empowerment processes vary significantly based on individual characteristics, with gender as a potentially influential factor (Balyer et al., 2017). Gelera-Capetillo (2014) found middle-aged, female, and married teachers demonstrated higher empowerment levels compared to other demographic groups. Similarly, Madriaga (2016) observed that female teachers with higher educational attainment showed greater empowerment capability exercise. These findings underscore gender as a critical variable in teacher empowerment assessment.

Existing literature on teacher empowerment measurement has primarily focused on scale development without explicit gender invariance consideration. Short and Rinehart (1992) developed the School Participant Empowerment Scale with six dimensions but did not extensively examine gender differences in scale functioning. While Hamadneh (2016) verified validity and reliability among teachers of gifted students, gender was only briefly addressed without comprehensive measurement invariance testing. Rangel et al. (2018) revalidated the scale for science and mathematics teachers, resulting in six revised factors, but again without thorough gender invariance examination.

Higher education contexts present particular gaps in understanding how teacher empowerment scales function across gender groups. Unlike primary and secondary education settings where some research exists, tertiary education involves unique organizational structures, professional expectations, and empowerment dynamics that may interact

differently with gender (Nessia, 2018). As noted by Carpenter (2015) and Al Salman and Hassan (2016), higher education settings often involve different forms of teamwork, collaboration, and professional autonomy that may be experienced differently across gender lines.

The relationship between gender and measurement instrument psychometric properties represents a critical consideration in scale development and validation. Kesebir et al. (2019) demonstrated that incorporating gender considerations in scale development leads to improved measurement outcomes. Pan et al. (2021) emphasized gender's essential role in scale development for accurately measuring responses and planning gender-specific interventions. Despite these insights, teacher empowerment scales specifically designed for tertiary education contexts have not been sufficiently examined for gender invariance.

This study addresses this significant gap by conducting comprehensive analysis of the Teacher Empowerment Scale's psychometric properties across gender groups in higher education. By employing Rasch analysis, which provides sophisticated methods for examining differential item functioning, this research offers insights into both overall measurement instrument quality and its fairness across gender groups. The findings contribute to developing more equitable assessment tools for teacher empowerment in tertiary educational settings, ultimately supporting more targeted and effective educational interventions aligned with global sustainability goals.

This study aimed to examine the measurement invariance and psychometric properties of the Teacher Empowerment Scale across gender groups using Rasch analysis. The research addressed the following objectives: To evaluate the overall Rasch fit statistics and reliability coefficients of the scale's three factors; To examine differential item functioning across gender groups; To assess the scale's measurement invariance and construct validity.

Methodological procedure

Research Design

The study employed Rasch analysis to examine the psychometric properties and measurement invariance of the Teacher Empowerment Scale (Panela, 2023). This analytical approach was selected to address the limitations of Classical Test Theory (CTT) and to provide a more robust examination of gender-based measurement invariance. The Rasch model is particularly appropriate for this investigation as it allows for sophisticated examination of how measurement instruments function across different demographic groups.

The analysis focused specifically on testing measurement invariance across teachers' gender. Measurement invariance refers to the assessment of construct equivalence between two or more groups, which is essential for making valid comparisons across gender categories (Bofah & Hannula, 2014). The process began with establishing a configural model in which all dimensions were freely estimated across gender groups. When a configural model demonstrates adequate

fit, researchers can reasonably assume that the same variables define each factor across gender groups. Following this, a weak invariance model was tested by constraining dimensional loadings to be equal across gender groups. When both configural and weak invariance models show support, additional constraints can be implemented to test for structural invariance. The presence of non-invariant structural models would suggest that the associations between underlying factors vary between male and female respondents (Bofah & Hannula, 2014).

The analysis concentrated on three established factors of the Teacher Empowerment Scale: fostering continuous improvement (comprising 51 items), teaching ownership and freedom (comprising 32 items), and work climate and conditions (comprising 3 items). This factor structure was previously established through exploratory and confirmatory factor analyses, providing a solid foundation for the current gender-based invariance testing.

Participants and Sampling

Data were collected from 968 higher education faculty members from state universities and colleges in Region VIII of the Philippines. The sample size exceeds the recommendations by Guadagnoli and Velicer (1988), who suggested that a minimum of 300-450 participants is required to achieve acceptable pattern comparability in factor analysis. The robust sample size strengthens the reliability of the findings and allows for meaningful subgroup comparisons by gender.

The sampling procedure employed a proportionate stratified random sampling technique to ensure adequate representation across institutions. This approach was selected because it provides better population coverage by giving the researcher control over subgroup representation (Fraenkel, Wallen & Hyun, 2011). Computer-generated selection was utilized to maintain randomness within these stratified groupings, thereby reducing potential selection bias.

Participants represented diverse academic disciplines, ranks, and years of teaching experience, enhancing the generalizability of findings within the higher education context. Due to pandemic-related restrictions during the data collection period, all survey administration was conducted online through secured digital platforms, ensuring participant safety while maintaining data integrity.

Data Collection

Data collection was conducted using the 86-item Teacher Empowerment Scale, which was developed through a rigorous process involving qualitative exploration, expert validation, and psychometric testing. The scale uses a 5-point Likert response format ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Prior to the main data collection, the instrument underwent pretesting with 325 college teachers from state universities and colleges in Region VIII, yielding a high reliability coefficient

(Cronbach's alpha = .947). This pretesting phase helped ensure that items were meaningful and significant to the target population before the full-scale administration.

For the main study, survey distribution occurred through official institutional channels, with formal permissions obtained from university administrators. Participants received detailed information about the study's purpose and were assured of confidentiality and anonymity. Digital consent forms were incorporated into the survey platform, requiring acknowledgment before participants could proceed to the questionnaire items.

To maximize response rates and minimize missing data, automated reminders were sent to participants who had not completed the survey, and partial responses were saved to allow completion across multiple sessions if needed. Data collection occurred over a three-month period to accommodate faculty teaching schedules and to ensure adequate representation across institutions.

Data Analysis

The inclusion of the Rasch Model is a response to the psychometric limitations of CTT which includes calibrations being sample and item dependent, incorrect assumed intervals for items with Likert scale and categories chosen are assumed the appropriate one (Boone, 2016). It is commonly used in IRT with the aim of describing the relationship between the level of ability of a person and the difficulty of the item. It a series of tests that consist of the following: polarity of the item, item fit statistics, characteristic curve of the item, differential item functioning, response category statistics, and the person-item map (Tran, Dorofeeva, & Loskutova, 2018).

Item polarity was evaluated using point-measure correlation coefficient (PTMEA CORR) which should display a high and positive item value of 0.3 to 0.8 which indicates that the items are working in the same direction to measure a single basic construct. An item outside of the said interval was recommended for removal (Boone, Yale, & Staver, 2014).

For item fit assessment, the mean-square (MNSQ) value should be between 0.5 to 2.0. Values less than the range do not bring efficiency in measurement building and any items more 2.0 was recommended for removal (Boone, 2016).

Item characteristic curve describes the relationship between the person's ability and probability of a correct response. Moreover, it also reflects the item difficulty which increases from left to right of the plot (Boone, 2016).

Differential item functioning (DIF) is used to evaluate the fit of each item based on a comparison of differences in proportion of correct responses between two groups of participants with equal ability. The study maintains that the probability of responses is not influenced by the

participants' gender. Therefore, differential item functioning analysis be gender grouping will be conducted to assess the characteristics of each item in this study.

The process measures DIF will use the following statistics: the Mantel chi-square statistic (Mantel), Standardized Liu-Agresti Cummulative Common Log-Odds Ratio (LOR Z), and Liu-Agresti Cumulative Common Log-Odds Ratio (L-A LOR). For the Mantel statistic, items with values above 3.84 (indicating a Type I error rate ≤0.05) will considered as presence of DIF. LOR Z values outside of the range from -1.96 to 1.96 represent evidence of DIF. L-A LOR values will be used to classify the size of the DIF: items with L-A LOR <0.53 are classified as class "A" because of the negligible amount of DIF, items with value between 0.53 and 0.74 belong to class "B" with moderate DIF, and items with a value of more 0.74 belong to class "C", containing high DIF. Items with class "C" will be excluded from the study (Boone, 2016).

For a subscale, response category statistics was conducted by combining all the items that belonged to that subscale into a single group. Categories statistics in each subscale required a gradual increase in level from category 1 (not empowered) to 5 (empowered) and no presence of reversal (Tran et al., 2018). On the person-item map, items were considered ideal when their distribution is sufficient to cover the distribution of a person (Boone et al., 2014).

Testing measurement or factor invariance across genders started with the aim of constructing a significant structural measurement model comparing teacher empowerment across gender. It will test the measurement invariance of the teacher empowerment scale between gender groups. To begin data analyses, a baseline model will be constructed that represents the teacher empowerment dimensions.

Multi-group Confirmatory Factor Analysis was used to analyze the invariance between gender groups. Measurement invariance can be analyzed with a four-step process which includes configural invariance, metric invariance, scale invariance, and strict invariance (Demir, 2017). For criteria to provide invariance, the hierarchic differences of model- data fit indices as well as the differences of model-data fit x2 statistics between the steps were considered using structural equation models (SEM). When the differences of the model-data fit indices were more than 0.01 and/or x2 statistics were statistically significant ($p < .05$), these findings were interpreted as a violation of invariance (Jorekog & Sorbom, 1996). Otherwise, it was decided that measurement invariance was provided across subgroups. In cases of some violations, the resources and causes of this violation were explained and discussed with a deep analysis based on the differences of the subgroups' model coefficients (Demir, 2017) using Akaike and Bayesian Information Criterion.

Ethical Considerations

This research adhered to rigorous ethical standards throughout its implementation. Prior to data collection, the study protocol received approval from the university's Institutional Research Ethics Committee. This approval ensured that the research methodology, data collection

procedures, and data management plans met established ethical guidelines for research involving human participants.

Informed consent was obtained from all participants before they completed the survey. The consent form detailed the study's purpose, procedures, potential risks and benefits, confidentiality measures, and the voluntary nature of participation. Participants were informed of their right to withdraw from the study at any point without penalty or consequence to their professional standing.

To protect participant confidentiality, all personally identifiable information was removed during data processing. Data were stored on password-protected servers with encrypted access limited to the research team. The reporting of results maintained anonymity by presenting only aggregated findings without identifying specific institutions or individuals.

Special consideration was given to the power dynamics inherent in educational institutions. The research team ensured that institutional administrators had no access to individual responses, protecting participants from potential professional repercussions. Additionally, the online administration method allowed participants to complete the survey in private settings, minimizing potential influence from colleagues or supervisors.

The study design also considered the principle of beneficence by ensuring that findings would contribute meaningful insights to the improvement of teacher empowerment in higher education, potentially benefiting participants and the broader educational community through enhanced understanding of gender-related factors in professional empowerment.

Results

The study included 968 teachers from State Universities and Colleges, utilizing online data collection due to COVID-19 pandemic health restrictions. Participants from baseline testing, in-depth interviews, pilot testing, and other phases were excluded from actual instrument administration. After scale administration, item homogeneity checking was conducted using Rasch analysis to verify that items reflect homogeneity in terms of gender—a trade-off between respondents' perceived factors affecting teacher empowerment.

Factor 1: Fostering Continuous Development

For factor 1 (fostering continuous development), 51 items were examined for fit statistics. Table 1 presents the overall fit analysis.

Table 1

Overall Rasch Fit Statistics and Reliability Coefficient for Factor 1

	Min	Max	Mean	SD
Infit MNSQ	0.87	1.31	1.00	0.09
Outfit MNSQ	0.85	1.22	0.99	0.09
Item Reliability		0.90		

The table presents the overall Rasch fit statistics and reliability coefficient of a set of items. The mean values of the Infit MNSQ and Outfit MNSQ are both close to 1.00, which indicates a good fit of the items to the Rasch model (Bond et al., 2015). The standard deviations of both Infit MNSQ and Outfit MNSQ are also relatively small, which further indicates that the fit is consistent across items (Linarce, 2011). The minimum and maximum values of the Infit MNSQ and Outfit MNSQ indicate that there are some items with slightly poorer fit than others, but overall, the fit is good. The item reliability coefficient is 0.90, which is considered to be high. This indicates that the set of items is reliable in measuring the construct of interest (Embretson et al., 2013). Overall, the table suggests that the set of items has a good fit to the Rasch model and is reliable in measuring the construct of interest.

According to Linacre (2011), an item should have infit and outfit mean squares of 1.0 to have a perfect fit, or between 0.5-1.5 to be productive for measurement. All items have infit and outfit mean squares inside the productive for measurement range. Further analysis highlights the differential item functioning (DIF) analysis for factor 1 according to gender. The analysis measures whether there are any differences in how males and females respond to the test items in Factor 1, which could indicate bias or unfairness in the test. Each row corresponds to a different test item, and the columns show various measures related to DIF. Overall, it appears that several items show significant DIF for gender, including item number 2, 10, 77, 79, and 86. The magnitude and direction of the DIF varies across items, with some items being associated more strongly with females and others with males.

Factor 2: Teaching Ownership and Freedom

For factor 2 (teaching ownership and freedom), 32 items were examined for fit statistics. Table 2 presents the overall fit analysis.

Table 2

Overall Rasch Fit Statistics and Reliability Coefficient for Factor 2

	Min	Max	Mean	SD
Infit MNSQ	0.60	1.18	1.00	0.10
Outfit MNSQ	0.61	1.20	0.96	0.11
Item Reliability			0.90	

The Infit MNSQ ranges from 0.60 to 1.18, with a mean of 1.00 and standard deviation of 0.10. The Outfit MNSQ ranges from 0.61 to 1.20, with a mean of .96 and standard deviation of 0.11. These values suggest that overall, the 32 items included in Factor 2 fit the Rasch model reasonably well. The mean values of the Infit MNSQ and Outfit MNSQ are both close to 1.00, which indicates a good fit of the items to the Rasch model (Bond et al., 2015). The standard

deviations of both Infit MNSQ and Outfit MNSQ are also relatively small, which further indicates that the fit is consistent across items (Linarce, 2011). The minimum and maximum values of the Infit MNSQ and Outfit MNSQ indicate that there are some items with slightly poorer fit than others, but overall, the fit is good. The item reliability coefficient is 0.90, which is considered to be high. This suggests that the items are measuring a common underlying construct (Embretson et al., 2013). Overall, based on the Rasch analysis results provided, Factor 2 appears to be a reliable and valid measure of the construct being assessed by the 32 items included in this factor.

According to Linacre (2011), an item should have infit and outfit mean squares of 1.0 to have a perfect fit, or between 0.5-1.5 to be productive for measurement. All items have infit and outfit mean squares inside the productive for measurement range. Further analysis highlights the differential item functioning (DIF) analysis for factor 2 according to gender. The analysis measures whether there are any differences in how males and females respond to the test items in Factor 2, which could indicate bias or unfairness in the test.

DIF occurs when the probability of responding to an item correctly differs between groups, in this case, males and females, even if they have the same underlying ability or construct being measured. Overall, it appears that several items show significant DIF for gender, including item number 62, 64, 71, and 72. The magnitude and direction of the DIF varies across items, with some items being associated more strongly with females and others with males. Another potential explanation for DIF is that the wording or content of the item is interpreted differently by males and females, leading to different probabilities of responding correctly.

Factor 3: Work Climate and Conditions

For factor 3 (work climate and conditions), 3 items were examined for fit statistics. Table 3 presents the overall fit analysis, while Table 4 shows item measures and item fit statistics

Table 3
Overall Rasch Fit Statistics and Reliability Coefficient for Factor 3

	Min	Max	Mean	SD
Infit MNSQ	0.94	0.99	0.97	0.02
Outfit MNSQ	0.75	0.81	0.79	0.03
Item Reliability			0.93	

The Infit MNSQ ranges from 0.94 to 0.99, with a mean of 0.97 and standard deviation of 0.02. The Outfit MNSQ ranges from 0.75 to 0.81, with a mean of 0.79 and standard deviation of 0.03. These values suggest that overall, the 3 items included in Factor 3 fit the Rasch model reasonably well. The mean values of the Infit MNSQ and Outfit MNSQ are both close to 1.00, which indicates a good fit of the items to the Rasch model (Bond et al., 2015). The standard deviations of both Infit MNSQ and Outfit MNSQ are also relatively small, which further indicates that the fit is consistent across items (Linarce, 2011). The minimum and maximum

values of the Infit MNSQ and Outfit MNSQ indicate that there are some items with slightly poorer fit than others, but overall, the fit is good. The item reliability coefficient is 0.93, which is considered to be high. This suggests that the items are measuring a common underlying construct (Embretson et al., 2013).

Overall, based on the Rasch analysis results provided, Factor 3 appears to be a reliable and valid measure of the construct being assessed by the 3 items included in this factor. According to Linacre (2011), an item should have infit and outfit mean squares of 1.0 to have a perfect fit, or between 0.5-1.5 to be productive for measurement. As seen on Table 27, all items have infit and outfit mean squares inside the productive for measurement range.

Table 4

Item Measures and Item Fit Statistics for Factor 3

Items	Item Measure	Standard Error	Infit MNSQ	Outfit MNSQ
66. No one supports my decision related to school obligation.	-0.13	0.08	0.98	0.75
67. The school has limited teaching resources.	0.40	0.07	0.99	0.81
68. There is no unity at work.	-0.27	0.08	0.94	0.80

Table 5 highlights the differential item functioning (DIF) analysis for factor 3 according to gender. The analysis measures whether there are any differences in how males and females respond to the test items in Factor 3, which could indicate bias or unfairness in the test. This table shows the results of a DIF analysis for Factor 3 by gender. Moreover, it appears that item number 66 has very little DIF, while item number 67 has moderate DIF in favor of females, and item number 68 has moderate DIF in favor of males.

Table 5

Differential Item Functioning for Factor 3 by Gender

Items	Female DIF Measure	Male DIF Measure	DIF Contrast	Joint SE	Welch T-value	p-value
66. No one supports my decision related to school obligation.	-0.07	-0.17	0.10	0.15	0.65	0.51
67. The school has limited teaching resources.	0.50	0.33	0.17	0.15	1.15	0.25
68. There is no unity at work.	-0.43	-0.15	-0.28	-0.15	1.81	0.07

Overall, the results suggest that while there is some DIF for Factor 3 by Gender for these three items, the differences are generally not statistically significant since all of the p-values were greater than the conventional threshold set at 0.05.

Discussion

Measurement Quality and Gender Invariance

This study provides extensive evidence for measurement invariance and psychometric quality of the Teacher Empowerment Scale across gender groups in higher education settings. The Rasch analysis results demonstrate that the scale functions effectively regardless of gender, while identifying specific areas where gender-based considerations may be relevant. The minimal differential item functioning across factors suggests that the instrument provides fair assessment of teacher empowerment constructs for both male and female educators, supporting its use in diverse educational contexts.

The strong item reliability coefficients across all factors (ranging from 0.90 to 0.93) indicate that the scale provides consistent measurement regardless of gender, meeting standards recommended by Embretson and Reise (2013) for high-quality psychometric instruments. These findings suggest that differences in empowerment scores between male and female faculty can be attributed to actual empowerment differences rather than measurement bias, providing confidence for administrators and researchers using this instrument.

Implications for Understanding Gender and Empowerment

The findings align with previous research by Madriaga (2016) and Gelera-Capetillo (2014) regarding gender differences in teacher empowerment, while demonstrating that these differences do not significantly impact the scale's measurement properties. The absence of significant DIF in the work climate and conditions factor supports observations by Boone (2016) and Linacre (2011) that organizational climate factors may be experienced more uniformly across demographic groups.

The presence of some DIF in factors 1 and 2, while minimal, provides insight into how certain aspects of empowerment may be perceived or experienced differently by male and female faculty. This finding does not invalidate the scale but rather highlights the importance of considering contextual factors when interpreting empowerment scores across gender groups. Educational administrators should be aware that while the scale provides fair measurement, underlying empowerment experiences may vary between male and female faculty members.

Contributions to Scale Development Theory

This study contributes significantly to addressing the gap identified by Kesebir et al. (2019) and Pan et al. (2021) regarding the need for gender considerations in scale development and validation. By establishing the gender neutrality of a comprehensive teacher empowerment measurement tool while acknowledging specific areas where gender-specific patterns emerge, this research advances the development of equitable assessment instruments for higher education settings.

The application of Rasch modeling in this context demonstrates the value of sophisticated psychometric approaches for examining measurement fairness. Unlike traditional approaches that might miss subtle forms of measurement bias, the Rasch model's focus on invariance properties provides more robust evidence for fair measurement across groups. This methodological contribution can inform future scale development efforts in educational contexts.

Relevance to Sustainable Development Goals

The establishment of a gender-invariant teacher empowerment scale directly supports multiple Sustainable Development Goals (SDGs). Most directly, this research advances SDG 5 (Gender Equality) by providing measurement tools that ensure fair assessment of empowerment across gender groups, reducing potential for gender bias in faculty evaluation and development processes. Additionally, the study supports SDG 4 (Quality Education) by contributing to improved faculty development frameworks that can enhance educational quality through better understanding of teacher empowerment dynamics.

The scale's capacity to fairly measure empowerment across gender groups enables educational institutions to design more targeted and effective interventions for faculty development. This capability is particularly important in higher education contexts where faculty empowerment directly impacts student learning outcomes and institutional effectiveness. By ensuring that measurement tools do not introduce gender bias, institutions can make more informed decisions about resource allocation and support programs.

Addressing Post-Pandemic Educational Challenges

The study's timing during and after the COVID-19 pandemic adds particular relevance to its findings. The pandemic significantly altered higher education landscapes, creating new challenges and opportunities for teacher empowerment. The establishment of a reliable, gender-invariant measurement tool provides institutions with the means to assess how these changes have affected faculty empowerment differently across gender groups.

The online data collection methodology, necessitated by pandemic restrictions, also demonstrates the scale's adaptability to contemporary educational contexts where digital platforms increasingly mediate professional interactions. This adaptability ensures the tool's continued relevance as educational institutions navigate hybrid and digital learning environments.

Limitations and Future Research Directions

While this study provides robust evidence for the Teacher Empowerment Scale's gender invariance, several limitations should be acknowledged. The sample was drawn from a specific geographic region (Region VIII of the Philippines), which may limit generalizability to other cultural contexts. Cultural factors influencing gender roles and professional empowerment may vary across different societies, potentially affecting how the scale functions in diverse international contexts.

The focus on gender as the primary demographic variable, while important, represents only one aspect of potential measurement bias. Future research should examine measurement invariance across other demographic variables such as age, years of experience, academic discipline, and educational attainment. Such investigations would provide a more comprehensive understanding of the scale's fairness across diverse faculty populations.

The cross-sectional design, while appropriate for establishing measurement invariance, does not capture potential changes in empowerment patterns over time. Longitudinal studies examining the stability of the scale's psychometric properties and gender invariance across different career stages would provide valuable insights into the dynamic nature of teacher empowerment in higher education.

Conclusion and Recommendation

This study provides comprehensive evidence for measurement invariance and psychometric quality of the Teacher Empowerment Scale across gender groups in higher education settings. Through rigorous Rasch analysis of 86 items across three factors, consistently good model fit and high reliability coefficients (ranging from 0.90 to 0.93) were found. The analysis revealed minimal differential item functioning across gender groups, with only a few items showing significant gender-based variations in factors 1 and 2, and none in factor 3.

The key contribution is that the Teacher Empowerment Scale provides fair and accurate assessment of teacher empowerment constructs for both male and female educators in higher education. The strong item reliability coefficients and good fit statistics across all factors confirm that the instrument meets standards recommended by Embretson and Reise (2013) and Linacre (2011) for high-quality psychometric instruments. These findings support the

scale's contribution to advancing SDG 5 (Gender Equality) and SDG 4 (Quality Education) through enhanced, bias-free faculty assessment frameworks.

The research addresses critical gaps in understanding gender considerations in educational measurement, providing empirical evidence that sophisticated psychometric approaches can detect and address potential measurement bias. The establishment of gender invariance enables educational institutions to confidently use this instrument for faculty development and institutional improvement initiatives.

Recommendations for Practice and Research

Based on these findings, the following recommendations are provided:

For Educational Practitioners:

1. Implement the Teacher Empowerment Scale with confidence across gender groups in higher education settings, while remaining attentive to specific items identified with differential functioning
2. Utilize the three-factor structure as a framework for designing targeted teacher empowerment interventions in higher education
3. Develop comparative studies examining empowerment patterns across different types of higher education institutions using this validated instrument
4. Explore relationships between teacher empowerment scores and educational outcomes to further establish the scale's practical utility

For Scale Development and Research:

1. Consider refinement of items showing significant DIF to enhance gender neutrality in future versions of the scale
2. Extend validation research to examine measurement invariance across additional demographic variables such as age, years of experience, and educational attainment
3. Conduct longitudinal studies to examine the stability of the scale's psychometric properties and gender invariance over time
4. Replicate this validation approach in diverse cultural contexts to establish international applicability

For Policy and Institutional Development:

1. Integrate findings into institutional policies for faculty development and empowerment assessment
2. Use the validated scale to establish baseline empowerment measures and track progress toward gender equality goals
3. Align empowerment measurement initiatives with institutional commitments to SDGs

4 and 5

4. Develop training programs for administrators on interpreting and acting upon empowerment assessment results

These recommendations collectively advance the goal of creating more equitable and effective higher education environments where all faculty members can experience and exercise professional empowerment regardless of gender, ultimately contributing to improved educational quality and outcomes.

REFERENCES

Al Salman, W., & Hassan, Z. (2016). Impact of effective teamwork on employee performance. International Journal of Accounting and Business Management, 4(1), 76-85.

Alvunger, D., & Wahlström, N. (2018). Research-based teacher education? Exploring the meaning potentials of Swedish teacher education. Teachers and Teaching: Theory and Practice, 24(4), 332-349.

Balkar, B. (2015). Defining an empowering school culture (ESC): Teacher perceptions. Issues in Educational Research, 25(3), 205-225.

Balyer, A., Ozcan, K., & Yildiz, A. (2017). Teacher empowerment: School administrators' roles. Eurasian Journal of Educational Research, 70, 1-18.

Bofah, E. A., & Hannula, M. S. (2014). Structural equation modelling: Testing for the factorial validity, replication and measurement invariance of students' views on mathematics. SAGE Research Methods Cases, 1-28.

Bond, T., & Fox, C. (2015). Applying the Rasch model: Fundamental measurement in the human sciences. Routledge.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? CBE Life Sciences Education, 15(4), rm4.

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. Springer.

Carpenter, D. (2015). School culture and leadership of professional learning communities. International Journal of Educational Management, 29(5), 682-694.

Chebet, R. (2014). Empowerment of teachers and performance of private secondary schools in Nomet County, Kenya. University of Nairobi: College of Humanities and Social Sciences.

Demir, E. (2017). Testing measurement invariance of the students' affective characteristics model across gender sub-groups. Educational Sciences: Theory and Practice, 17(1), 47-62.

Embretson, S. E., & Reise, S. P. (2013). Item response theory. Psychology Press.

Fraenkel, J., Wallen, N., & Hyun, H. (2011). How to design and evaluate research in education. McGraw-Hill Education.

Gelera-Capetillo, N. (2014). Empowerment, morale and performance of elementary grade teachers in the Division of Calbayog City. Northwest Samar State University.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. Psychological Bulletin, 103(2), 265-275.

Hamadneh, B. M. (2016). Level of empowerment of teachers of gifted students in schools of King Abdullah II for Excellence, Jordan. Journal of Education and Human Development, 5(1), 262-269.

Joreskog, K. G., & Sorbom, D. (1996). LISREL 8: User's reference guide. Scientific Software International.

Kesebir, S., Lee, S., Elliot, A. J., & Pillutla, M. M. (2019). Lay beliefs about competition: Scale development and gender differences. Motivation and Emotion, 43(5), 719-739.

Linacre, J. M. (2011). A user's guide to WINSTEPS. Winsteps.com.

Madriaga, C. A. (2016). Lay leaders' empowerment in the Salesian schools. International Journal of Education and Research, 4(5), 393-408.

Nessia, R. (2018). Project TEACH: Educator empowerment program launched. Philippine Information Agency.

Oracion, C. C. (2015). Teacher leadership in public schools in the Philippines. Institute of Education, University of London.

Pan, L., Lu, L., & Zhang, T. (2021). Destination gender: Scale development and cross-cultural validation. Tourism Management, 83, 104225.

Panela, C. (2023). Development and validation of teacher empowerment scale: Examining factor structure and Rasch model fit in higher education. [Unpublished doctoral dissertation]. University of the Philippines.

Rangel, D., Monter, M., Olyai, A., Lowry, S., & Sloan, T. (2018). Revalidation of the School Participant Empowerment Scale among science and mathematics teachers. Journal of Psychoeducational Assessment, 36(3), 303-314.

Short, P. M., & Rinehart, J. S. (1992). School Participant Empowerment Scale: Assessment of level of empowerment within the school environment. Educational and Psychological Measurement, 52(4), 951-960.

Simon, R. (2013). Performance behavior and morale of elementary teachers of Tinambacan District of Calbayog City. Northwest Samar State University.

Tran, V. D., Dorofeeva, V. V., & Loskutova, E. E. (2018). Development and validation of a scale to measure the quality of patient medication counseling using Rasch model. Pharmacy Practice, 16(4), 1327.