POLARIS GLOBAL JOURNAL OF SCHOLARLY RESEARCH AND TRENDS

Research Article

Development and Validation of Gender- Sensitive and Inclusive Educational Resources Evaluation Tool

https://doi.org/10.58429/pgjsrt.v5n1a225

Jackielee A. Anacleto[1], Jick C. Balinario[2], Mary Ann T. Arceno[3]
State University of Northern Negros, Philippines
Email: janacleto@sunn.edu.ph, jcbalinario@sunn.edu.ph, marceno@sunn.edu.ph

VOLUME 5 |NO. 1 |2026 ISSUE

ABSTRACT: Several legal frameworks mandate the integration of gender, diversity, equity, and inclusion in curricula and educational resources. Despite these mandates, a significant gap exists between policy and practice. In reality, discrimination, biases, and stereotypes are still found in many educational resources. This disparity not only contradicts the legal and ethical imperatives but also prevents the development of gender-responsive, sensitive, and inclusive education. The lack of a measurement tool to evaluate the gender-sensitivity and inclusivity of educational resources is a challenge in bridging the gap. This study aims to develop and validate a comprehensive measurement tool to evaluate gender-sensitivity and inclusivity of educational resources, employing a Research and Development (R&D) approach. The study systematically establishes the instrument through stages of conceptualization, development, expert review, and pilot testing. A sample of 80 educational professionals with at least five years of teaching experience participated in the validation procedures, which included content validity assessment through expert and construct validity via Confirmatory Factor Analysis (CFA) using Jamovi software. The results indicated that the developed instrument demonstrates high content validity, all items attained a Content Validity Index of 1, and exhibited strong construct validity supported by fit indices in CFA. Reliability analysis yielded a Cronbach's alpha coefficient within the excellent range, confirming the tool's consistency. The study concludes that this tool can serve as a standard rubric for policymakers, educators, and curriculum developers to have a clear and objective mechanism in identifying biases, measuring inclusivity, and ensuring compliance with legal and moral obligations.

KEYWORDS
Gender Sensitive, Education, Evaluation, Jamovi, CFA.

INTRODUCTION

The research study, focusing on the development and validation of a standardized evaluation instrument, is not merely an academic endeavor but a necessary mechanism for state compliance and institutional accountability in the Philippine Higher Education Institution (HEI) sector. This initiative is directly anchored in both international commitments and national legal mandates aimed at achieving substantive gender equality.

In educational institutions, gender stereotypes, prejudices, discrimination, gender-based violence (GBV), and other types of gender subordination and oppression are still common (Reyes et al., 2021). According to recent research, there is still GBV in schools (e.g. Valencia, M.I.C. & Reyes, Z. Q., 2023 & Sorbring et al., 2022).

Furthermore, others argue that biased curricula are the reason why gender disparity still exists (Amin, S., Girard, C., & Calabrò. G.S. (2022). Therefore, achieving gender equality requires not just equal access to education but also the elimination of all types of discrimination against women and girls in curricula, pedagogy, language, and instructional materials.

While CHED CMO 1, s. 2015, mandates GAD mainstreaming there is no standardized tool that goes beyond general policy adherence to specific, trustworthy content analysis of all educational resources (e.g., textbooks, modules, e-learning resources), It is still challenging to accurately quantify and scale GAD implementation across the country without such a standardized and tested tool. The university has taken steps to assist instructors in creating inclusive and gender-sensitive teaching materials. To evaluate how inclusive and gender-sensitive the training materials are, a standard instrument must be developed. In order to ensure compliance with national gender mainstreaming mandates, the researcher developed and validated a Gender-Sensitive and Inclusive Educational Resources Evaluation Tool for use in the university, with a primary focus on the objective assessment of language, content, and images in instructional materials.

Statement of the Problem/Objectives of the Study

The study aims to develop and validate a measurement tool for gender-sensitive and inclusive educational resources. Specifically, it answers the following questions:

1. Identify the legal foundations for gender-sensitive and inclusive educational resources.
2. Develop a measurement tool for gender-sensitive and inclusive educational resources.
3. Validate the developed measurement tool to expert in the field of Gender and development.
4. Determine the construct validity and reliability of the developed measurement tool for gender-sensitive and inclusive educational resources.

METHODS

a. Research Design

The research employs a rigorous Research and Development methodology to create and validate a measurement tool specifically designed for gender-sensitive and inclusive educational resources, ensuring the final product is both valid and reliable through systematic and scientific procedures. The utilization of research instruments in education and social science research is paramount, dictating the research design through data collection and the retrieval of accurate information, which is then reported in the study's results (Hamzah et al., 2014). This robust approach ensures the tool meets stringent standards of accuracy and consistency (Toma & Lederman, 2020). The process of validation is a critical component of ensuring the quality and credibility of research instruments (Drost, 2011). The instrument is subjected to validation by a pool of experts to ensure content validity (C. Mercado, 2020). The R&D methodology provides a structured framework for not only creating the measurement tool but also rigorously testing its effectiveness within the intended educational context (Scott et al., 2019).

b. Sampling Method

This study underscores the importance of adhering to established guidelines for sample size determination in quantitative research, especially when validating measurement tools, and it contributes to ensuring the reliability and validity of research outcomes in educational contexts. The determination is based on the widely recognized rule of thumb of having at least 5 observations for each independent variable, given that the "Gender-Sensitive and inclusive Educational Resource Measurement Tool" includes 16 independent variables. In the context of the current study, a sample size of 80 respondents is deemed appropriate. This approach aligns with recommendations from prominent scholars in the field of research methodology (White, 2022). The selection of an adequate sample size is not merely a numerical consideration; it is deeply intertwined with the statistical power of the study, the generalizability of the findings, and the overall rigor of the research process.

c. Data Collection

The researcher gathered data based on the stages of the study. In conceptualization stage, the researcher gathered and conducted a group discussion to teachers who have training on gender and development which is the bases in the formulation of a framework or model, followed by identifying and defining the construct of the proposed framework or model.

In development stage, the researcher established an initial version of the instrument refine and extended by the research team after the conduct of group discussions with the selected

teachers and then presented to them and instructional materials developers, and revise the instrument according to their feedback.

In the expert review stage, the researcher sought 5 experts in the field of Gender and Development and Curriculum experts to review and validate the develop measurement tool for gender-sensitive and inclusive educational resources and further revise the develop instrument according to their feedback.

In the pilot study, the researcher conducted to 80 faculty in one state university. Each of them had at least 5 years of experience as a teacher. This stage is also a stage of testing the construct validity and confirming the reliability of the instrument for use in full- scale studies. The participants were first introduced to the instrument and its purpose and then requested to fill it individually and provide any feedback that they find relevant.

d. **Data Analysis**

To answer SOP 3, As a strategy to validate the content of instrument construction, content validity index (CVI) was the method that used by the researchers. Likewise, Confirmatory Factor Analysis (CFA) was used to verify whether a set of observed variables accurately reflects the underlying theoretical constructs they are intended to measure. The process began by clearly defining the latent variables and hypothesizing how observed indicators relate to these constructs, grounded in existing theory or prior research. Researchers then collected and prepared data, ensuring it meets necessary assumptions such as normality and adequacy for factor analysis. Using jamovi, the hypothesized model was specified and estimated, typically employing methods like Maximum Likelihood estimation. The model's fit is assessed through various indices, including the Chi-square test, RMSEA, CFI, TLI, and SRMR, to determine how well the proposed structure aligns with the observed data. If the model does not fit well, modifications were made based on theoretical justification and statistical indicators, such as modification indices, to improve alignment. This rigorous approach ensures that the measurement model is both statistically sound and theoretically meaningful, providing a robust framework for understanding complex constructs (Statistic Solutions, n.d.; Number Analytics, 2025). To answer SOP 4, the standard of reliability follows the interpretation of Cronbach's alpha coefficient range (<0.6 (weak); 0.6–0.7 (moderate); 0.7–0.8 (good); 0.8–0.9 (very good); >0.9 (excellent)).

RESULTS/FINDINGS

Legal and Policy Frameworks **Supporting the Development of Gender-Neutral Educational Resources Tool**

Table 1. Legal foundations for gender-sensitive and inclusive educational resources.

Statements	Legal Bases	References
LANGUAGE:	1.Universal Declaration of Human Rights (UDHR) 2. Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) 3. Sustainable Development Goals (SDGs) SDG 4 (Quality Education) SDG 5 (Gender Equality) 4. 1987 Philippine Constitution 5. Republic Act No. 9710 (Magna Carta of Women) 6. CMO No. 1, series of 2015 7. Department of Education (DepEd) Order No. 32, s. 2017 (Gender-Responsive Basic Education Policy):	1. United Nations. (1948). Universal Declaration of Human Rights 2. United Nations. (1979). Convention on the Elimination of All Forms of Discrimination against Women. . 3. United Nations. (n.d.). Sustainable Development Goals. Retrieved from https://sdgs.un.org/goals 4. Constitution of the Republic of the Philippines. (1987) 5. Republic Act No. 9710. (2009, August 14). An Act providing for the Magna Carta of Women\ 6. Commission on Higher Education. (2015). CMO No. 1, series of 2015: Revised policies and guidelines on student affairs and services 7. Department of Education. (2017). Gender-Responsive Basic Education Policy (DepEd Order No. 32, s. 2017)
1. Addresses all genders as persons of equal value.		
2. Addresses all genders as persons of equal dignity.		
3. Makes all genders visible or represented.		
4. Talks of intersectionality with respect to diversity.		
5. Use of gender neutral words.		
CONTENT:		
1. Mentions all genders significantly.		
2. Describes gender equity and equality among all genders in gaining access and control.		
3. Describes gender equity and equality among all genders in participation.		
4. Describes gender equity and equality among all genders in sharing responsibilities and benefits.		
5. Depicts in a balanced manner all genders as to psychological traits and competence.		
6. Mentions all genders, with regards to their intersectionionalities, with respect to diversity.		
Images/Illustrations:		
1. Portrays all genders with equal value.		
2. Portrays all genders with equal dignity.		
3. Shows varied or neutral colors.		
4. Depicts integration of all genders.		
5. Portrays all genders, with regards to their intersectionionalities, with respect to diversity.		

Table 1 presents a comprehensive overview of the legal and policy frameworks supporting the development of gender-neutral educational resources tool. These frameworks span international declarations, national legislation, and sectoral policies that collectively emphasize equity, inclusion, and diversity in educational content and delivery.

Language Considerations. The Universal Declaration of Human Rights (UDHR) (United Nations, 1948), CEDAW (United Nations, 1979), and Sustainable Development Goals (SDGs)

(United Nations, n.d.) lay foundational principles for non-discrimination and the equal dignity of all individuals, regardless of gender. These instruments encourage the use of inclusive language that makes all genders visible, equally represented, and respected. The mention of intersectionality reflects a modern understanding of gender as intersecting with other identity markers, such as ethnicity, class, and ability, affirming the SDG principle of "leaving no one behind." Nationally, the 1987 Philippine Constitution and Republic Act No. 9710 (Magna Carta of Women) reinforce these global commitments by mandating the use of language that affirms gender equality and avoids stereotypes. The Commission on Higher Education (CHED) CMO No. 1, series of 2015, and the DepEd Order No. 32, s. 2017, go further by institutionalizing gender-responsive approaches in curricular and extracurricular activities in both basic and higher education. These policies ensure that language used in educational materials is inclusive, avoids gender bias, and acknowledges the diversity of learners.

Content. The content-related items in Table 1 are directly aligned with the legal mandates to ensure gender equality in access, participation, and benefits. The emphasis on equity in control and participation mirrors the principles outlined in CEDAW and the Magna Carta of Women, which seek to eliminate discrimination in all forms and empower all genders through equal opportunity. The inclusion of intersectionality as a content criterion is consistent with both the SDGs and DepEd's gender-responsive education policy, which recognize that social justice in education must address multiple, overlapping systems of marginalization**.**

Images and Illustrations. Visual elements in educational materials play a significant role in shaping perceptions. The legal foundations cited in Table 1 promote equal representation and positive portrayals of all genders. DepEd Order No. 32 and CHED's CMO No. 1 advocate for non-stereotypical imagery, the use of varied or neutral colors, and visuals that reflect integration and diversity. These policies are intended to combat the perpetuation of gender roles and promote positive role modeling across gender identities.

Validate the developed measurement tool with an expert in Gender and development.

Development Research

The development of instruments was based on correct and scientific processes and procedures. Based on the instrument development procedure, after the items of the instrument have been arranged, the next stage was the validity of the construct through a qualitative study by five experts that aims to review the instruments that have been prepared. After the qualitative expert review, the next stage was assessment through an expert panel quantitatively. The result of the panelist assessment was analyzed using the Content Validity Index. Based on the analysis using Lawse content validation, it indicated that all items of the instrument have Context Validity Index

values of 1 as validated by a 5-expert panel, which means that all the items are appropriate or can be used without improvement.

Testing Validity and Calculating Reliability of the Gender-Sensitive Educational Resource Measurement Tool

In this research, the standardization of the "Gender-Sensitive and Inclusive Educational Resource Measurement Tool" uses confirmatory factor analysis. The analysis factor in this research uses Confirmatory Factor Analysis (CFA) with the help of jamovi. The measurement model was based on the goodness of fit criteria, which are used to test the fit of the theoretical model with empirical data.

Measurement Model (Not Corrected)

Table 2. Assessment of Not Corrected Standardized factor loadings

Factor	Indicator	Estimate	SE	Z	p	Stand. Estimate
Language	L1	0.0987	0.0193	5.12	<.001	0.559
	L2	0.2178	0.1319	1.65	0.099	0.192
	L3	0.2308	0.0227	10.18	<.001	0.937
	L4	0.0864	0.0149	5.79	<.001	0.612
	L5	0.1426	0.0231	6.18	<.001	0.650
Content	C1	0.1946	0.0239	8.13	<.001	0.771
	C2	0.2120	0.0207	10.24	<.001	0.895
	C3	0.2238	0.0218	10.28	<.001	0.897
	C4	0.1738	0.0219	7.93	<.001	0.758
	C5	0.1860	0.0252	7.39	<.001	0.723
	C6	0.2080	0.0178	11.70	<.001	0.964
Image/Illustration	I1	0.1900	0.0152	12.47	<.001	1.000
	I2	0.1628	0.0153	10.61	<.001	0.912
	I3	0.1484	0.0198	7.51	<.001	0.726
	I4	0.1048	0.0240	4.36	<.001	0.466
	I5	0.1048	0.0203	5.17	<.001	0.537

Based on the results of the measurement model analysis in Table 2, in factor 1 (Language) it shows item no. 1, 3, 4, and 5 have a loading factor (λ)> 0.5, which means that the items are valid and fit to be used for data collection in accordance with the opinion of Hendryadi and Suryani (2014). Table 2 shows that all manifest variables had a value of α = 0.01, which means that the relationship between manifest variables and factors or indicators is significant. The Language factor showed a strong loading for L3 (0.937), indicating a robust relationship with the latent construct. However, L2 had a low loading (0.192) and was not statistically significant, suggesting it may not be a reliable indicator for this factor. Likewise, The Content factor exhibited consistently high loadings across all indicators, with C6 showed the highest loading (0.964), indicating a strong association with the underlying construct. Lastly, the Image/Illustration factor demonstrates strong loadings,

particularly for I1 (1.000) and I2 (0.912), suggesting these indicators were highly representative of the construct. However, I4 had a lower loading (0.466), indicating a weaker relationship. High factor loadings should be at least .5 and ideally .7 or higher (Hair et.al., 2010). The loading should be statistically significant as indicative of a strong relationship between the observed variables and their underlying latent constructs. So, the factor with less than 0.50 suggests a need for further examination or potential revision of this indicator. Furthermore, Table 1 shows that there is no change, in which items of the statement have a factor load of > 0.5, which means that statements were valid.

Table 3. Convergent Validity of the Instrument

Factor	Stand. Estimate	Stand. Estimate²	Measurement Error (1-SE)	Composite reliability CR (>0.5)	Average Variance Extracted (>0.5)
C1	0.559	0.312481	0.441		
C2	0.192	0.036864	0.808		
C3	0.937	0.877969	0.063	0.809347	0.404872
C4	0.612	0.374544	0.388		
C5	0.65	0.4225	0.35		
Total	**2.95**	**2.024358**	**2.05**		
L1	0.771	0.594441	0.229		
L2	0.895	0.801025	0.105		
L3	0.897	0.804609	0.103		
L4	0.758	0.574564	0.242	0.961952	0.704444
L5	0.723	0.522729	0.277		
L6	0.964	0.929296	0.036		
Total	**5.008**	**4.226664**	**0.992**		
I/I1	1	1	0		
I/I 2	0.912	0.831744	0.088		
I/I 3	0.726	0.527076	0.274		
I/I 4	0.466	0.217156	0.534	0.907019	0.572869
I/I5	0.537	0.288369	0.463		
Total	**3.641**	**2.864345**	**1.359**		

Table 3 shows the Language factor exhibits a CR above the acceptable threshold of 0.7, indicating good internal consistency. However, the AVE is below the recommended value of 0.5, suggesting that the construct may not adequately capture the variance of its indicators. Notably, one indicator has a low loading of 0.192, which could be contributing to the reduced AVE. Furthermore, the Content factor demonstrates both high CR and AVE values, indicating strong internal consistency and that the construct explains a substantial portion of the variance in its indicators. Furthermore, The Image/Illustration factor also shows acceptable CR and AVE values, supporting the convergent validity of this construct. However, the lower loading of 0.466 for one indicator suggests that this item may not align well with the underlying construct. The analysis of Table 3

indicates that the Content and Image/Illustration factors exhibit satisfactory convergent validity, as evidenced by their high CR and AVE values. The Language factor, while demonstrating acceptable internal consistency (CR), falls short in terms of AVE, suggesting that some indicators may not effectively capture the intended construct. Further refinement of the Language factor's indicators may enhance its convergent validity.

Table 4: Model Fit

Test for Exact Fit			Fit Measure			
χ^2	df	p	CFI	TLI	SRMR	RMSEA
463	101	<.001	0.692	0.634	0.163	0.212

Table 4 shows the model fit of the instrument. The chi-square value is 463 with 101 degrees of freedom, and a p-value less than .001. A significant chi-square indicates a discrepancy between the observed and expected covariance matrices, suggesting a poor model fit. However, it's important to note that the chi-square test is sensitive to sample size, and significant results are common in large samples. Comparative Fit Index (CFI) value is 0.692. Values above 0.90 are generally considered indicative of acceptable fit, with values above 0.95 indicating good fit (Hu & Bentler, 1999). Therefore, a CFI of 0.692 suggests a poor fit. Tucker-Lewis Index (TLI) value is 0.634. Similar to the CFI, values above 0.90 are desirable, indicating that the model does not fit the data well. Standardized Root Mean Square Residual (SRMR) value is 0.163. Values less than 0.08 are generally considered acceptable (Hu & Bentler, 1999). Thus, an SRMR of 0.163 indicates a poor fit. Root Mean Square Error of Approximation (RMSEA) value is 0.212. Values less than 0.06 indicate good fit, and values up to 0.08 represent reasonable errors of approximation in the population (Browne & Cudeck, 1993). A value of 0.212 is substantially higher, suggesting a poor fit. Hu and Bentler (1999) recommend that a combination of CFI ≥ 0.95, TLI ≥ 0.95, SRMR ≤ 0.08, and RMSEA ≤ 0.06 indicates a good model fit. The values reported in Table 3 fall short of these thresholds, suggesting that the model does not adequately fit the data.

Furthermore, Kline (2015) emphasizes the importance of considering multiple fit indices when evaluating model fit, as reliance on a single index can be misleading. Given that all reported indices in Table 4 indicate poor fit, it is advisable to re-express the model, possibly by revising the factor structure or considering alternative models. The fit indices in Table 4 suggest that the current model does not adequately fit the data. All indices fall outside the recommended thresholds for acceptable model fit, indicating the need for model respecification. Future analyses should consider alternative model structures and re-evaluate the fit indices to achieve a model that better represents the underlying data structure.

Table 5. Assessment of Discriminant validity

Factor	Language	Content	Image/Illustration
Language	**0.63**	0.732	0.186
Content	0.732	**0.83**	0.476
Image/Illustration	0.186	0.476	**0.75**

Note. Values in the diagonal are the square root of the AVE's

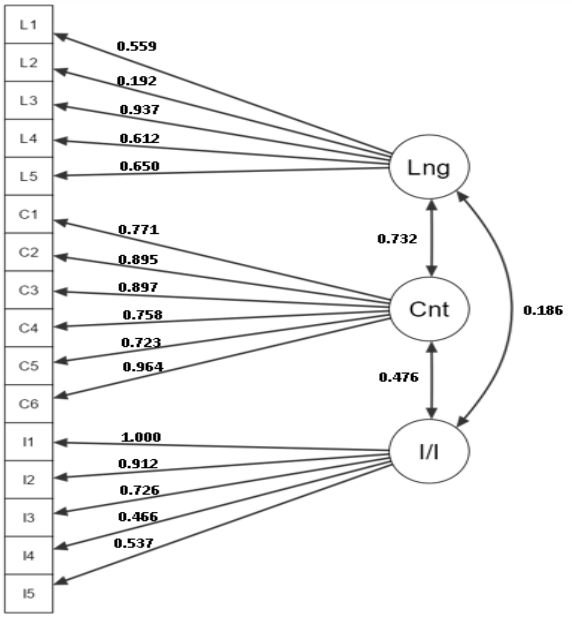

Figure 2 : Path Diagram (Not Corrected)

Table 5 presents the assessment of discriminant validity among three constructs: Language, Content, and Image/Illustration. The diagonal elements represent the square roots of the Average Variance Extracted (AVE) for each construct, while the off-diagonal elements denote the inter-construct correlations. For Language and Content, the correlation between Language and Content is 0.732, which is higher than the square root of AVE for Language (0.63). This suggests a lack of discriminant validity between these two constructs. Likewise, Language and Image/Illustration the correlation is 0.186, which is lower than the square roots of AVE for both constructs (Language: 0.63; Image/Illustration: 0.75), indicating adequate discriminant validity. Finally, Content and Image/Illustration the correlation is 0.476, which is lower than the square roots of AVE for both constructs (Content: 0.83; Image/Illustration: 0.75), suggesting acceptable discriminant validity.

The interpretation was based on Hair et.al (2010) which establish discriminant validity, the square root of AVE should be higher than the related correlations. The analysis reveals that while the constructs of Language and Image/Illustration, as well as Content and Image/Illustration, demonstrate adequate discriminant validity, the high correlation between Language and Content suggests a lack of discriminant validity between these two constructs. This finding implies that the items measuring language and content may not be sufficiently distinct and could capture overlapping aspects of the constructs. It is recommended to revisit the operational definitions and measurement items for these constructs to enhance their discriminant validity

Measurement Model

The standardized estimates in Table 6 confirmed the strength and reliability of the measurement model. A loading value above 0.70 is generally considered acceptable for construct validity (Hair et al., 2022).

Table 6. Assessment of Corrected Standardized factor loading

Factor	Indicator	Estimate	SE	Z	p	Stand. Estimate
Language	**L3**	0.226	0.0256	8.81	<.001	0.917
	L5	0.144	0.0237	6.08	<.001	0.656
Content	**C1**	0.195	0.0239	8.12	<.001	0.771
	C2	0.212	0.0207	10.25	<.001	0.895
	C3	0.224	0.0218	10.27	<.001	0.896
	C4	0.174	0.0219	7.92	<.001	0.758
	C5	0.186	0.0251	7.42	<.001	0.725
	C6	0.208	0.0178	11.70	<.001	0.964
Image/Illustration	**I1**	0.192	0.0154	12.47	<.001	1.008
	I2	0.162	0.0156	10.33	<.001	0.905
	I3	0.147	0.0199	7.42	<.001	0.720

Nearly all indicators meet or exceed this threshold, with Language L3 (0.917), Content C6 (0.964), and Image I1 (1.008) showing very strong relationships with their latent constructs. Even the lowest value, L5 (0.656), remains within an acceptable range for social science research (Brown, 2015). These results imply that the indicators effectively represent their respective latent constructs. For example, the Content factor demonstrates strong internal consistency across all six indicators (ranging from 0.725 to 0.964), suggesting a comprehensive and coherent measurement of content quality. Similarly, Image/Illustration indicators showed high reliability (all >0.70), underscoring the clarity and visual support provided by illustrations in educational or media content. The Z-values (all > 6) further confirmed the statistical significance of each indicator's contribution to the model, indicating that none of the measured variables were redundant or weak (Schumacker & Lomax, 2021). The p-values (< .001) confirm this across all paths, suggesting strong model validity.

Table 7 provides evidence for the convergent validity of the instrument using two key psychometric indicators: Composite Reliability (CR) and Average Variance Extracted (AVE). These metrics were essential in structural equation modeling and confirmatory factor analysis that assessed how well a set of indicators represents a latent construct (Hair et al., 2022; Fornell & Larcker, 1981).All three latent constructs Language (CR = 0.854), Content (CR = 0.962), and Image/Illustration (CR = 0.950) exceeded the minimum CR threshold of 0.70, indicating high internal consistency (Raykov & Marcoulides, 2019). A CR value above 0.80 was considered strong, and all three constructs surpass this benchmark, suggesting that the instrument consistently measures each latent factor (Zhao et al., 2021). Furthermore, The AVE values were also robust:

Language (0.636), Content (0.705), and Image (0.784) all exceed the 0.50 threshold, demonstrating that over 50% of the variance is explained by the latent variable rather than by error (Fornell & Larcker, 1981; Byrne, 2016). High AVE values further confirm that the indicators strongly reflect their underlying construct and that the measurement model possesses good convergent validity. Notably, the Image/Illustration construct demonstrates exceptionally strong convergent validity (AVE = 0.784), which aligns with growing research emphasizing the role of visual elements in enhancing comprehension and engagement in educational materials (Mayer, 2021; Clark & Mayer, 2016).

Table 7. Corrected Convergent Validity of the Instrument

Factor	Stand. Estimate	Stand. Estimate²	Measurement Error (1-SE)	Composite reliability CR (>0.5)	Average Variance Extracted (>0.5)
L3	0.917	0.840889	0.083	0.852826	0.635613
L5	0.656	0.430336	0.344		
Total	**1.573**	**1.271225**	**0.427**		
C1	0.771	0.594441	0.229		
C2	0.895	0.801025	0.105	0.962003	0.704628
C3	0.896	0.802816	0.104		
C4	0.758	0.574564	0.242		
C5	0.725	0.525625	0.275		
C6	0.964	0.929296	0.036		
Total	**5.009**	**4.227767**	**0.991**		
I/I 1	1.008	1.016064	-0.008		
I/I 2	0.905	0.819025	0.095	0.949724	0.784496
I/I 3	0.72	0.5184	0.28		
Total	**2.633**	**2.353489**	**0.367**		

The findings in Table 7 affirm that the instrument was both reliable and valid for assessing the constructs of Language, Content, and Image/Illustration. This validation was crucial for ensuring meaningful results in subsequent structural modeling or practical application in curriculum and instructional design. Educators and researchers can confidently utilize this tool for evaluating learning materials, ensuring alignment between instructional design and cognitive processing principles (Mayer, 2021). The strength of the Content construct, in particular, implies that the instrument was highly effective in gauging the quality and depth of instructional content—a critical factor for learner achievement and engagement (Zhang et al., 2023).

Table 8: Corrected Model Fit

Test for Exact Fit			Fit Measure			
χ^2	df	p	CFI	TLI	SRMR	RMSEA
133	41	<.001	0.91	0.90	0.0727	0.167

Table 8 presents the model fit indices used to evaluate the overall goodness-of-fit of the confirmatory factor analysis (CFA) model. The results included the chi-square statistic (χ^2), degrees of freedom (df), p-value, and several widely accepted fit indices such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Standardized Root Mean Square Residual (SRMR), and Root Mean Square Error of Approximation (RMSEA). These indices collectively provide insight into how well the hypothesized model represents the observed data (Hair et al., 2022; Kline, 2016).

The fit indices reported in Table 8 provided insight into how well the proposed measurement model aligns with the observed data. The chi-square test of model fit was statistically significant, $\chi^2(41) = 133$, $p < .001$, indicating a lack of exact fit. However, due to the known sensitivity of the chi-square test to sample size, researchers often rely on alternative fit indices for a more accurate assessment (Kline, 2016; Schumacker & Lomax, 2016). The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were both at 0.91 and 0.90, respectively. These values exceed the minimum recommended threshold of 0.90, suggesting an acceptable model fit (Hu & Bentler, 1999; Hair et al., 2022). While values above 0.95 are preferred for a "good" fit, the reported indices suggest the model is adequately specified, particularly for early-stage or exploratory measurement models (Byrne, 2016). The Standardized Root Mean Square Residual (SRMR) was 0.0727, which falls below the 0.08 cutoff, further indicating adequate fit (Weston & Gore, 2006). However, the Root Mean Square Error of Approximation (RMSEA) was 0.167, which was considerably above the recommended threshold of 0.06 for good fit or even the 0.08 cutoff for acceptable fit (Browne & Cudeck, 1993; MacCallum et al., 1996). This suggests poor fit, possibly due to model misspecification, measurement error, or the need for model respecification (e.g., adding correlated errors or removing poorly loading indicators).

Table 9 presents the discriminant validity results for the latent constructs "Language," "Content," and "Image/Illustration" in the measurement model. The Fornell–Larcker criterion was used to assess discriminant validity, where the square root of the AVE for each construct (shown on the diagonal) must be greater than its correlations with any other construct (Fornell & Larcker, 1981). The square root of the AVE for Language is 0.79, which is higher than its correlation with Content (0.744) and Image/Illustration (0.150). The square root of the AVE for Content is 0.83, which was also greater than its correlations with Language (0.744) and Image/Illustration (0.475). The square root of the AVE for Image/Illustration is 0.88, which exceeds its correlations with Language (0.150) and Content (0.475). These results indicate that each construct shares more variance with its own indicators than with other constructs, thereby demonstrating adequate discriminant validity (Hair et al., 2022; Kline, 2016).

Table 9. Corrected of Discriminant validity

Factor	Language	Content	Image/Illustration
Language	**0.79**	0.744	0.150
Content	0.744	**0.83**	0.475
Image/Illustration	0.150	0.475	**0.88**

Note. Values in the diagonal are the square root of the AVE's

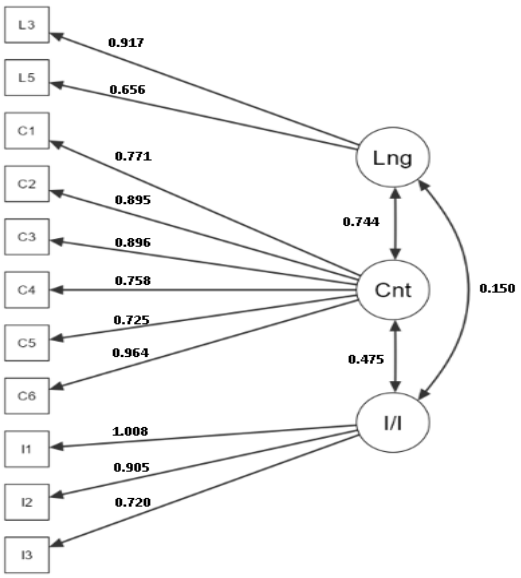

Figure 3: *Path Diagram (Corrected)*

The confirmation of discriminant validity implied that the constructs "Language," "Content," and "Image/Illustration" were empirically distinct from one another. This supports the theoretical structure of the instrument and strengthens the overall construct validity of the measurement model (Campbell & Fiske, 1959; Henseler et al., 2015). However, it was worth noting that the correlation between Language and Content (r = 0.744) approaches the threshold for concern. Although discriminant validity is technically established, the high inter-construct correlation suggests potential conceptual overlap and may indicate the need for closer review of item content in future studies to ensure distinctiveness (Farrell, 2010). Ensuring strong discriminant validity is particularly important for structural equation modeling (SEM), as failure to do so may result in biased estimates and misleading conclusions in the structural model (Hair et al., 2022; Brown, 2015).

Reliability Analysis

Table 10: Reliability Analysis

Factors	Not Corrected Cronbach's α	Corrected Cronbach's α
Language	0.308	0.748
Content	0.930	0.930
Image/Illustration	0.854	0.900
Total	0.786	0.915

Table 10 presents the internal consistency reliability of the instrument's subscales, as measured by Cronbach's alpha (α) before and after correction. Cronbach's alpha was a measure of internal consistency, or how closely related a set of items are as a group. An alpha of ≥ .70 was

considered acceptable, ≥ .80 was good, and ≥ .90 was excellent (George & Mallery, 2019; Nunnally & Bernstein, 1994).

The language initially had a low reliability score (α = .308), indicating poor internal consistency. After correction—likely through item deletion or refinement—its reliability improved to an acceptable level (α = .748). Furthermore, Content exhibited excellent reliability both before and after correction (α = .930), suggesting very high item cohesion within this factor. While, image/Illustration also showed good to excellent reliability, improving from α = .854 to α = .900 after correction. The total scale reliability improved significantly from α = .786 (acceptable) to α = .915 (excellent), demonstrating strong consistency across the instrument after adjustments.The corrected reliability coefficients indicate that the revised measurement instrument possesses adequate to excellent internal consistency across all factors. The significant improvement in the Language factor suggests that the original items may have been either ambiguous or poorly aligned with the construct, necessitating item refinement—a common step in scale development (DeVellis, 2017). High reliability in the Content and Image/Illustration factors confirms that these sections are measuring their intended constructs consistently, which supports the construct validity of the instrument (Tavakol & Dennick, 2011). Additionally, the overall instrument's corrected alpha of .915 suggests that it can be confidently used for further analyses such as structural equation modeling or outcome evaluation (Hair et al., 2022). Reliable instruments are essential in ensuring that results are replicable and generalizable, reducing measurement error and enhancing the credibility of findings (Kline, 2016).

CONCLUSION/IMPLICATION OF THE STUDY AND RECOMMENDATION

Based on the findings the following conclusions can be drawn for each objective:

1. Through a rigorous process encompassing systematic item formulation, expert review, and thorough psychometric evaluation, a comprehensive and validated measurement tool for assessing gender-neutral educational resources has been successfully developed. The instrument, encompassing key constructs like language, content, and visual elements, demonstrates strong content validity, construct validity, and reliability, making it suitable for practical application in evaluating educational materials.
2. The development and validation of this measurement tool for gender-neutral educational resources are implicitly grounded in legal and ethical principles, emphasizing adherence to scientific standards and institutional compliance in educational research. The instrument's proven validity and reliability provide a valuable resource for educators and policymakers seeking to ensure that educational

materials promote gender neutrality and comply with relevant standards, ultimately fostering a more equitable and inclusive learning environment.

Based on the objectives and findings of the study, the following recommendations are proposed:

1. To ensure widespread impact, educators, curriculum developers, and educational evaluators should actively adopt the validated measurement tool for assessing gender neutrality in educational resources. Accompanying this adoption, educational institutions should provide comprehensive training workshops for educators, content creators, and evaluators on effectively utilizing the tool and emphasizing the importance of gender sensitivity and inclusivity in educational resource development.
2. To maximize the long-term effectiveness and relevance of the measurement tool, ongoing research and user feedback should be actively encouraged to refine and update the instrument, ensuring it remains aligned with evolving perspectives on gender and educational equity. Furthermore, advocacy efforts should focus on integrating this measurement approach into official educational standards and accreditation processes to systematically promote gender-inclusive education.

REFERENCES

Brown, T. A. (2015). Confirmatory factor analysis for applied research (2nd ed.). Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Testing structural equation models (pp. 136–162). Sage.

Byrne, B. M. (2016). Structural equation modeling with AMOS: Basic concepts, applications, and programming (3rd ed.). Routledge.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81–105. https://doi.org/10.1037/h0046016

Clark, R. C., & Mayer, R. E. (2016). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning (4th ed.). Wiley.

Commission on Higher Education. (2015). CMO No. 1, series of 2015: Revised policies and guidelines on student affairs and services. https://ched.gov.ph/cmo-1-s-2015/

Constitution of the Republic of the Philippines. (1987). https://www.officialgazette.gov.ph/constitutions/1987-constitution/

Daniel, J., & Siddiqui, N. (2023). Special issue: Gender equality and education. *Review of Education*, *11*(3). https://doi.org/10.1002/rev3.3456

De Mesa, R. Y. H., Marfori, J. R. A., Fabian, N. M. C., Camiling-Alfonso, R., Javelosa, M. A. U., Bernal-Sundiang, N., Dans, L. F., Calderon, Y. T., Celeste, J. A., Sanchez, J. T., Rey, M. P., Galingana, C. L. T., Paterno, R. P. P., Catabui, J. T., Lopez, J. F. E., Aquino, M. R. N., & Dans, A. M. L. (2023). Experiences from the Philippine grassroots: impact of strengthening primary care systems on health worker satisfaction and intention to stay. BMC health services research, 23(1), 117. https://doi.org/10.1186/s12913-022-08799-1

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. Trials, 15, 267. https://doi.org/10.1186/1745-6215-15-267

Department of Education. (2017). DepEd Order No. 32, s. 2017: Gender-Responsive Basic Education Policy. https://www.deped.gov.ph/2017/07/28/do-32-s-2017-gender-responsive-basic-education-policy/

DeVellis, R. F. (2017). Scale development: Theory and applications (4th ed.). SAGE Publications.

Eustaquio, P. C., Dela Cruz, J. D. M., Araña, Y., Rosos, B., Rosadiño, J. D. T., Pagtakhan, R. G., Regencia, Z. J. G., & Baja, E. S. (2023). Prevalence of and factors associated with the use of gender-affirming hormonal therapy outside the reference regimen among transgender people in a community-led clinic in Metro Manila, Philippines: a retrospective cross-sectional study. BMJ open, 13(9), e072252. https://doi.org/10.1136/bmjopen-2023-072252

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). Journal of Business Research, 63(3), 324–327. https://doi.org/10.1016/j.jbusres.2009.05.003

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.2307/3151312

George, D., & Mallery, P. (2019). IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference (16th ed.). Routledge.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). Multivariate data analysis (9th ed.). Cengage Learning.

Hansen, J. J., & Gissel, S. T. (2020). Learning platform pedagogic : Learning platforms as a pedagogical framework, pedagogical planning tool and time and place of learning. *Research Portal Denmark*, 20.

Hansen, J. J., Gissel, S. T., & Gissel, S. T. (2020). Discourses of Danish as a subject on learning platforms: didactic analysis of courses for Danish L1 teaching. *Research Portal Denmark*, 227.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. Electronic Journal of Business Research Methods, 6(1), 53–60.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55.Wikipedia

Initiative on GAD, 2015. https://pcw.gov.ph/the-national-gender-and-resource-program/ Retrieved 02/24/2024

Huang, P., & Liu, X. (2024). Challenging gender stereotypes: representations of gender through social interactions in English learning textbooks. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03293-x

Joo, Y. J., Lim, K. Y., & Kim, E. K. (2022). The impact of content quality on learner satisfaction in online education. Educational Technology Research and Development, 70(3), 811–829. https://doi.org/10.1007/s11423-021-09989-0

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Meng, Y., Xu, W., Liu, Z., & Yu, Z. (2024). Scientometric analyses of digital inequity in education: problems and solutions. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03480-w

Metaxa, D., Wang, K., Landay, J. A., & Hancock, J. T. (2018). *Gender-Inclusive Design*. https://doi.org/10.1145/3173574.3174188

Mayer, R. E. (2021). Multimedia learning (3rd ed.). Cambridge University Press.

Number Analytics. (2025). 7 Essential Steps in Confirmatory Factor Analysis for Data. Retrieved from https://www.numberanalytics.com/blog/confirmatory-factor-analysis-steps

Nunes, I., Moreira, A., & Araújo, J. (2022). GIRE: Gender-Inclusive Requirements Engineering. *Data & Knowledge Engineering*, *143*, 102108. https://doi.org/10.1016/j.datak.2022.102108

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

O'Brien, K., Petra, V., Lal, D., Kwai, K., McDonald, M., Wallace, J., Jeanmonod, C., & Jeanmonod, R. (2022). Gender coding in job advertisements for academic, non-academic, and leadership positions in emergency medicine. The American journal of emergency medicine, 55, 6–10. https://doi.org/10.1016/j.ajem.2022.02.023

Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. PloS one, 14(5), e0216241. https://doi.org/10.1371/journal.pone.0216241

Peng, J., Shen, W. Q., Rao, J., & Lin, J. (2025). Automated Bias Assessment in AI-Generated Educational Content Using CEAT Framework. In *Communications in computer and information science* (p. 352). Springer Science+Business Media. https://doi.org/10.1007/978-3-031-99264-3_44

Qiao, C., Chen, Y., Guo, Q., & Yu, Y. (2024). Understanding science data literacy: a conceptual framework and assessment tool for college students majoring in STEM. *International Journal of STEM Education*, *11*(1). https://doi.org/10.1186/s40594-024-00484-5

Raykov, T., & Marcoulides, G. A. (2019). Introduction to psychometric theory (2nd ed.). Routledge.

Republic Act No. 9710. (2009, August 14). The Magna Carta of Women. https://pcw.gov.ph/republic-act-9710/

Schmiedel, T.; vom Brocke, J.; Recker, J.(2014). Development and Validation of an Instrument to Measure Organizational Cultures'Support of Business Process Management. Inf. Manag. 2014, 51, 43–56. [CrossRef]

Schumacker, R. E., & Lomax, R. G. (2021). A beginner's guide to structural equation modeling (5th ed.). Routledge.

Statistic Solutions. (n.d.). Confirmatory Factor Analysis (CFA): A Detailed Overview. Retrieved from https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/confirmatory-factor-analysis/

Shah, S. M. (2025). *Gender Bias in Artificial Intelligence: Empowering Women Through Digital Literacy*. https://doi.org/10.70389/pjai.1000088

Son, M., & Ha, M. (2024). Development of a digital literacy measurement tool for middle and high school students in the context of scientific practice. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12999-z

Spagnolo, C., & Nicchiotti, B. (2023). Interpreting gender gap issues in standardized tests: definition and application of a theoretical tool. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1303041

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

United Nations. (1948). Universal Declaration of Human Rights. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (n.d.). Sustainable Development Goals. https://sdgs.un.org/goals

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. The Counseling Psychologist, 34(5), 719–751. https://doi.org/10.1177/0011000006286345

Yañez, A. G. B., Alonso-Fernández, C., & Fernández-Manjón, B. (2023). Systematic literature review of digital resources to educate on gender equality. *Education and Information Technologies*, *28*(8), 10639. https://doi.org/10.1007/s10639-022-11574-8

Yu, S., Hsueh, Y.-L., Chih-Yuan Sun, J., & Liu, H.-Z. (2021). Developing an intelligent virtual reality interactive system based on the ADDIE model for learning pour-over coffee brewing. *Computers and Education. Artificial Intelligence*, *2*, 100030. https://doi.org/10.1016/j.caeai.2021.100030

Zabaniotou, A., Boukamel, O., & Tsirogianni, A. (2021). Network assessment: Design of a framework and indicators for monitoring and self-assessment of a customized gender equality plan in

the Mediterranean Engineering Education context. *Evaluation and Program Planning*, *87*, 101932. https://doi.org/10.1016/j.evalprogplan.2021.101932

Zhang, J., Kim, S., & Zhao, R. (2023). Content richness and user engagement in digital learning. Computers & Education, 200, 104821. https://doi.org/10.1016/j.compedu.2023.104821

Zhao, X., Lynch, J. G., & Chen, Q. (2021). Reconsidering convergent and discriminant validity in consumer research. Journal of Consumer Research, 48(2), 455–472. https://doi.org/10.1093/jcr/ucab011