

Research Article

Ensuring the reliability of content validity form for education, linguistics, and social sciences

Nedie Sevilla ^{1,*} , Michael Estremera ¹ , Amancio Cantoria ¹ , Vener Jeresano ² , April Joy Acosta ³ ¹ Dr. Emilio B. Espinosa Sr. Memorial State College of Agriculture and Technology, Philippines² Bicol University, Philippines³ Iloilo State University of Fisheries and Technology, Dingle Campus, Philippines

* Correspondence: nedie.sevilla@debesmscat.edu.ph

<https://doi.org/10.59652/hnes7027>

Abstract: An effective and reliable content validity form encompasses numerous important components, such as quantitative measures and qualitative notes. It must capture the concept or field being explored, offering a distinct description and scope. This paper employs a mixed-methods study approach that combines a quantitative reliability investigation with qualitative professional authentication by subjecting the newly developed content validity form to face validation conducted by specialists in the fields of education, linguistics, social sciences, philosophy, and anthropology at the site of interest. Leveraging the online Cronbach's alpha calculator, Kruskal-Wallis test, Shapiro-Wilk normality test, and Fleiss' kappa to determine the agreement reliability of collected data, this paper confirms that content validity forms are indispensable tools for guaranteeing research excellence across the fields of education, dialectology, social sciences, philosophy, and anthropology. The recognized components – lucidity, bearing, inclusiveness, and alignment – determine both applicability and discipline-particular variations. Reliability examination vouches that well-developed forms produce reliable outcomes, while professional authentication highlights the position of iterative improvement and multidisciplinary partnership. These discoveries offer a robust basis for developing effective and dependable instruments, supporting laborious research and evaluation practices in varied academic arenas. Researchers now recommend that the newly developed tool be used within the identified disciplines to maximize its benefits in field research.

Keywords: content validity; content validity form; education; linguistics; reliability

Received: April 6, 2026

Accepted: June 15, 2026

Published: June 27, 2026



Copyright: © 2022 by the authors.

Submitted for open access publication

under the terms and conditions of the

Creative Commons Attribution (CC BY)

license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An ideal and adoptable content validity form is indispensable for guaranteeing that dimensions in the developed or adopted instruments and questionnaires, along with instruction, linguistics, philosophy, and anthropology-related studies, are precisely representing the paradigms researchers envision to explore (Masuwai et al. 2024; Mollman et al. 2024). Correspondingly, content validity covers the degree to which a test or calculation widely captures all pertinent features of the subject matter, warranting that no sensitive elements are misplaced and that extraneous parts are omitted. Along with educational inquiries, content validity is commonly assessed by subject matter specialists who identify whether the questionnaire parts or research tool itself sufficiently capture the area of interest (Shrotryia & Dhanda, 2019). In linguistics, philosophy, and anthropology, there are perceived to be greater challenges, as these cognitive domains typically investigate abstract and socially nuanced or context-dependent paradigms that require comprehensive operationalization and validation.

Basically, a valid and reliable content validity form comprises numerous important components. It captures the concept or field being explored, offering a distinct description and scope (Govindasamy et al. 2024; Patrick et al. 2011). Moreover, it also presents the items or questions involved in the questionnaire or instrument, each of which must have been straightforwardly connected to the research construct. The form should solicit experts' comments on the weight, clearness, and representativeness of each entry. Essentially, expert raters are usually requested to assess items on a Likert-type measure, signifying whether each

element is indispensable, useful but not vital, or extraneous. The form may similarly have space for qualitative remarks, permitting specialists to recommend enhancements or add-ons. In the context of the study, the form computes a content validity index (CVI) or content validity ratio (CVR), which calculates correspondingly the degree of agreement among specialists concerning the tool content and scope (Gupta et al. 2025; Tay & Jebb, 2018).

Further, the trustworthiness feature of a content validity form relies on numerous aspects (Boateng et al., 2018; Carmines and Zeller, 1979). The selection of specialists is crucial; they must possess the necessary expertise and comprehension in the field under investigation. The lucidity and uniformity of the directions given to experts likewise influence consistency, as vague directions can result in unreliable evaluations. The instrument itself must have been well-developed and easy to respond to, curtailing the danger of mistakes or misinterpretations. In addition, the procedure of collecting expert evaluations – whether via simple agreement, numerical indices, or more intricate procedures – must be clear and replicable (Elangovan & Sundaravel, 2021).

In spite of the extensive use of content validity forms, some research gaps persist. Initially, there is a nonexistence of homogenous guiding principles for developing and gauging content validity forms in the different fields, predominantly in disciplines such as philosophy and anthropology, where theories are commonly intangible and context-reliant (Dalawi et al., 2023; Masuwai et al. 2024; Mollman et al. 2024). Besides, a maximum number of existing studies emphasize quantitative procedures of content validity, like CVI or CVR; nevertheless, there is a dearth of investigation of qualitative approaches for evaluating content validity, like thematic scrutiny of expert remarks. There is likewise a necessity for more explorations on the consistency of content validity procedures, together with studies that scrutinize the reliability of professional assessments over time and across diverse frameworks. Notably, there is a perceived gap in the works concerning the incorporation of content validity procedures with other kinds of validity, like construct and criterion validity, predominantly in interdisciplinary research that blends instruction, dialectology, philosophy, and anthropology similarly (Uddin, 2025).

Thus, from the highlighted premises, this article is grounded on the following research questions: (1) What are the components of an ideal and adoptable content validity form for education, linguistics, social sciences, philosophical and anthropological studies? (2) What is the reliability measurement of the content validity form for education, linguistics, social sciences, philosophical & anthropological studies? (3) How reliable is the content validity form for education, linguistics, social sciences, philosophical & anthropological studies?

2. Materials and Methods

2.1. Research Design

This study considers a mixed-methods study approach, particularly combining quantitative reliability investigation and qualitative professional authentication (Gierus et al. 2025). This was undertaken by subjecting the innovated content validity form to face validation of specialists in the disciplines of education, linguistics, social sciences, philosophy, and anthropology in the site of interest. Thereafter, data obtained was subjected to reliability analysis to obtain the quantitative data. Correspondingly, the form also included open-ended questions to collect qualitative data, which might not be captured by the quantitative part of the content validation.

Basically, this methodological piece of work centered on the creation of instrument and the psychometric assessment of a content validity form that can potentially be used in the social sciences, linguistics, and education field. The domains and items on the form were initially and qualitatively reviewed and improved for clarity and relevance by a panel of experts from aforementioned fields. In order to evaluate agreement and thoroughly verify reliability, their ratings had been quantitatively analyzed using indices like the CVI, CVR, and modified kappa. The results, in fact, guided the researchers for iterative revision to create a solid, broadly applicable content validity form.

2.2. Criteria Selection of Validators

Table 1 presents the qualified validators based on the selection criteria purposely set by the researcher. These criteria have to do with fields of expertise, 15 years of experience in government service, and having a Google Scholar account with publications in reputable journals. As a matter of fact, the mathematics validator is expected to look into the accuracy of the scale and other quantitative aspects of the innovated tool (Aslan et al. 2025).

Table 1. Validators' profile and remark.

Discipline	Length of service	Google Scholar account	Quantitative remark	Descriptive rating
Mathematics	22	https://scholar.google.com/citations?user=xq_oZJ4AAAAJ&hl=en	9.8	Highly valid
English	28	https://scholar.google.com/citations?user=4MmHYT0AAAAJ&hl=en	9.5	Highly valid
Social Science and Anthropology	20	https://scholar.google.com/citations?user=qUYk7kMAAAAJ&hl=en	9.8	Highly valid

The English expert, on the other hand, examines both the qualitative aspects and the grammatical and diction accuracy of the form. Essentially, the social science expert examines the suitability of the content validity form in the fields of anthropology, philosophy, and other allied fields, ensuring that the form will resonate with the existing body of knowledge locally and internationally.

In detail, validators were purposefully chosen based on their academic standing in the social sciences, languages, and education, as well as their methodological experience and disciplinary knowledge. In effect, they must have a good publication or research record in their discipline, at least a master's degree (ideally a PhD), and must have eminent experience in curriculum design, assessment, or instrument creation. To guarantee both depth and diversity of judgment, each disciplinary cluster had a minimum of at least 1 to 2 specialists. Moreover, at least five years of professional or instructional experience, as well as knowledge of content validity or relevant psychometric principles, are additional requirements.

2.3. Data Collection and Collation

The main business of the study is to highlight the reliability of the innovated form (Appendices A and B) to contribute to the target fields. This collective tool was distributed to the target and qualified validators in the second week of October 2025 through a Google Forms survey and was collated using Microsoft Excel for accuracy and precision purposes.

Accordingly, the content validity form was administered online or by email to the chosen expert validators in the social sciences, education, and languages. Thereafter, using a standardized Likert-type scale, validators independently assessed each item representativeness, relevance, and clarity and offer qualitative feedback for improvement. Then, the completed forms were reviewed for completeness and returned within the predetermined time frame. After that, the quantitative ratings will be entered into a spreadsheet to calculate CVR, CVI, and associated indices, and the qualitative comments coded and compiled thematically to direct the content validity form iterative revision.

2.4. Statistical Treatment of Data

This paper leverages numerous statistical tests to satiate the quantitative component and similarly to quantify the reliability component of the innovated tool. This involves basic statistics such as mean, standard deviation, median, skewness, kurtosis, and sum of squares. Statistical tools are needed to calculate the relationships between the form components, as validated by the participants. As for the reliability measurement, the standard alpha of the three validators was obtained using the Cronbach's alpha calculator. The Kolmogorov-Smirnov and Shapiro-Wilk normality tests were undertaken to determine the normality of data collected. Most significantly, the Fleiss' kappa test was done to determine inter-rater agreement reliability.

As a matter of fact, the statistical analysis concentrated on measuring inter-expert agreement and content validity at the item and scale levels. This includes the relevance and clarity assessments of experts to calculate the item-level and scale-level content validity indices (I-CVI, S-CVI/Ave, and S-CVI/UA). Further, to ascertain which items satisfy minimum essentiality requirements, CVR was computed. Besides, to evaluate consistency among validators, chance-corrected agreement had been analyzed using a modified kappa coefficient or related indices. Ultimately, to assess the stability and robustness of the content validity form, ratings from the social sciences, linguistics, and education were summarized using descriptive statistics.

2.5. Data Analysis

The quantitative and qualitative components of this study lead to the use of quantitative

and thematic-content analyses techniques to highlight the captured data. Verbatim analysis of validators was performed using themes and codes to substantiate the quantitative data. In fact, in order to assess the content validity form dependability and utility in the fields of education, linguistics, and social sciences, data analysis combined quantitative indices and qualitative comments. This aims primarily to assess item retention, revision, or deletion. Quantitative item ratings had likewise been descriptively examined (means, standard deviations, and % agreement) and utilized to calculate I-CVI, S-CVI, CVR, and modified kappa. To assess stability, patterns of agreement had also been compared between discipline groupings. To explain statistical trends and direct the improvement of item wording, domain structure, and form usage instructions, qualitative feedback from validators were thematically coded

3. Results

The following section presents the quantitative and qualitative findings regarding the development and validation of the content validity form. The results examine the validity, reliability, and cross-disciplinary applicability of the instrument through descriptive statistics, normality and agreement tests, expert evaluations, and thematic analysis of qualitative feedback. Collectively, these findings provide evidence of the instrument robustness, consistency, and potential utility for researchers in education, linguistics, social sciences, and related disciplines.

Table 2 highlights the components of the content validity, which should similarly be reflected in the questionnaires to be developed by researchers and scientists in the fields of linguistics, education, anthropology, and other allied disciplines, ensuring that their instrument will practically capture what it intends to obtain

Table 2. Components of content validity form.

Content validation criteria	Components
<i>I. Quantitative measure</i>	
The questionnaire is brief and limited in scope that target participants would feel easy to accomplish without having to require much of their precious time answering.	Admissibility and consideration
The questionnaire bears clarity and conciseness of instructions that potential respondents will not exert much effort in trying to grasp them.	Clarity and conciseness
The questionnaire is geared towards collection of accurate data and aligned with appropriate research design; otherwise, it will most likely collect flawed and inaccurate data which may heavily affect research outcome.	Accuracy and appropriateness
The questionnaire considers sensitivity and being compassionate to the respondents and does not have humiliating and personal questions which may make target users to be uneasy and apprehensive while answering the questions.	Sensitivity and being compassionate
The developed questionnaire bears the element of being time- oriented which implies that it will not take too much effort accomplishing and responding.	Time-oriented
The questionnaire is within the ambit and scope of the topic being explored. It will gather appropriate and essential data from the target participants.	Suitability
The questionnaire promotes flexibility and adaptableness to certain unlooked-for conditions that respondents can still be able to respond.	Flexibility
The developed questionnaire has the element of relatability to the target audience making the included questions easy to answer.	Relatability
The questionnaire undergoes proofreading for possible lexical and grammatical errors.	Precision
The questionnaire considers legibility and does not have too much colors which may distract the target users.	Legibility
The questionnaire fosters sensitivity to gender orientation, age, ethnicity, and religion.	Gender-oriented
The questionnaire offers ample options where respondents can comfortably choose the most relevant answer from the given choices.	Relevance and comfort
The statements included in the questionnaire consider the principles of truthfulness and reliability.	Truthfulness and dependability
The questionnaire could be administered to a wider audience and varied groups to achieve higher validity of research findings.	Validity and wide-ranging

The questionnaire is well-matched with other ensuing data such as transcriptions and datasets resulting to a substantial raw data capturing completely the research site.	Substantiality
The questionnaire does not demand the intended users to calculate higher statistics just to satisfy what is asked for in the questionnaire.	Simplicity and being considerate
The questionnaire observes alignment with research objectives, research questions and hypothesis.	Alignment
The questionnaire does not force the respondents to give personal information thereby respecting the privacy option of the target participants.	Fairness
The questionnaire takes into consideration originality of information which may redound to a better grasp of existing policy, issues and concerns, and/or program.	Originality and novelty
The questionnaire wisely considers statements that will resonate both in local and international perspectives contributing to the body of knowledge.	Relatability with wider audiences
<i>II. Qualitative notes</i>	
How do you assess the overall validity and quality of the instrument? What are your further comments for improvement?	Flexibility, quality, and accuracy
Please give your honest remarks about the alignment of the questionnaire/ survey form to the research questions, hypotheses/assumptions.	Research questions, hypothesis/assumption alignment
How will the questionnaire data contribute to the existing body of knowledge and international audiences? Please indicate the parts needing refinement.	Novelty and relatability

In fact, some of the noted essential components of questionnaires have to do with admissibility and consideration, clarity and conciseness, accuracy and appropriateness, sensitivity and being compassionate, time-orientedness, suitability, relatability, precision, legibility, gender-orientedness, relevance and comfort, truthfulness and dependability, validity and wide-rangingness, substantiality, simplicity and being considerate, fairness, originality, flexibility and accuracy, hypothesis and assumption alignment, and most essentially, novelty components of the instrument (Plengkham et al. 2025).

Based on the data reflected in Figure 1, the mean value of 9.91 represents the collective CVI score of the form, denotative of a robust content validity. Conversely, a standard deviation value of 0.15 is comparatively small, demonstrating a tight cluster of data points around the mean. In the context of the study, this number could imply that the innovated form bears a high level of precision, with experts' ratings being closely aligned. It also suggests consistency among experts' opinions, indicating strong agreement on the form content validity.

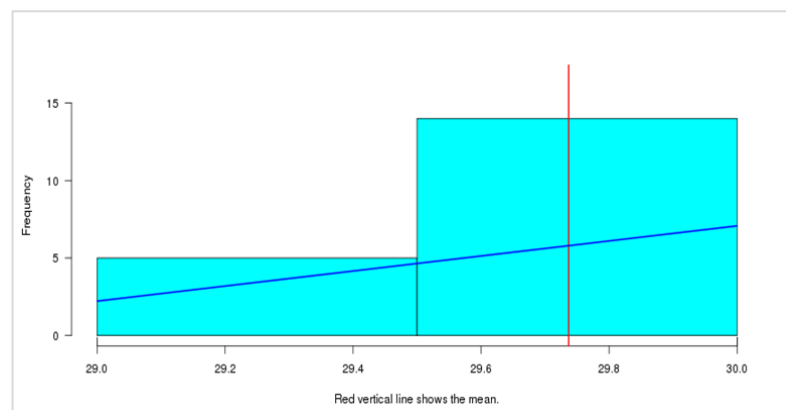


Figure 1. Distribution of expert content validity ratings for the content validation form.

Moreover, as for the 9.92 trimmed value, it likely shows that the mean score of the content validity form is 9.92 after removing a certain percentage of the highest and lowest scores (trimming). This communicates a more robust estimate of the content validity, reducing the impact of extreme scores. In fact, the high trimmed mean score indicates strong content validity, as experts rated the form items as relevant and important. To supplement, a median absolute deviation value of 9.67 indicates the median distance between each data point and the median of the dataset, implying a moderate variability in experts' ratings, suggesting slight disagreement among experts. Be that as it may, it is less likely to be affected by outliers, indicating the form content validity is generally strong with noted slim variation.

Additionally, a skew value of -1.06 indicates the distribution of scores is negatively skewed, implying that most experts rated the content validity form high, with a few low scores pulling the mean down. This implies the content validity form is generally well-regarded, but some experts had concerns or disagreed with certain components. On the other hand, a kurtosis value of 0.03 indicates the distribution of scores is mesokurtic, indicating that the distribution is close to normal, with no extreme tails (neither too flat nor too peaked). This directly implies the data is comparatively well-behaved, with no major deviations from normality, supportive of the reliability of the content validity form results.

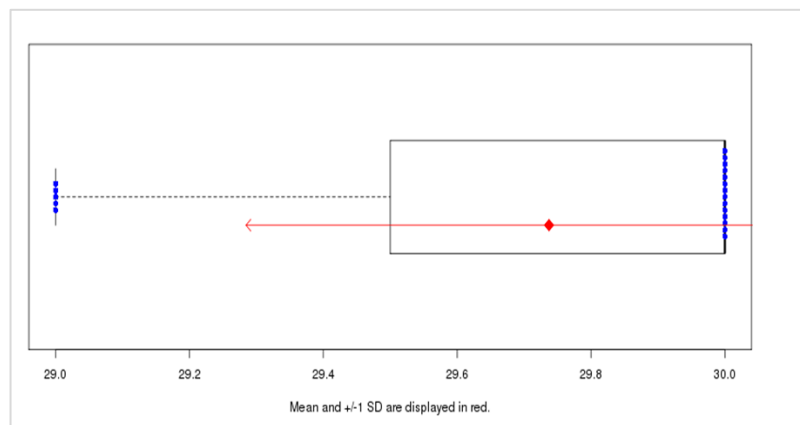


Figure 2. Box plot with individual data points.

The content validity form ratings from experts in the fields of education, linguistics, and social sciences had been visually summarized using a box plot with individual data points. In fact, figure 2 indicates that the overall spread and possible outliers, while the box shows the median and interquartile range of scores. To display the underlying distribution and clustering of judgments, individual expert assessments superimposed as jittered points. In order to help the interpretation of the quantitative content validity indices, the graphic makes it easier to compare the central tendency, variability, and consistency of ratings among disciplinary groups.

To establish if the variances exist among the experts' validation, collected data was subjected to a non-parametric test using the Kruskal-Wallis test. Thus, the computed $H = 0.6885$ value indicates no significant difference (likely $p > 0.05$), confirming that groups may have similar distributions or median scores. As a matter of fact, the content validity form performance is consistent across groups. This suggests the form reliability and components are largely constant across different fields or clusters.

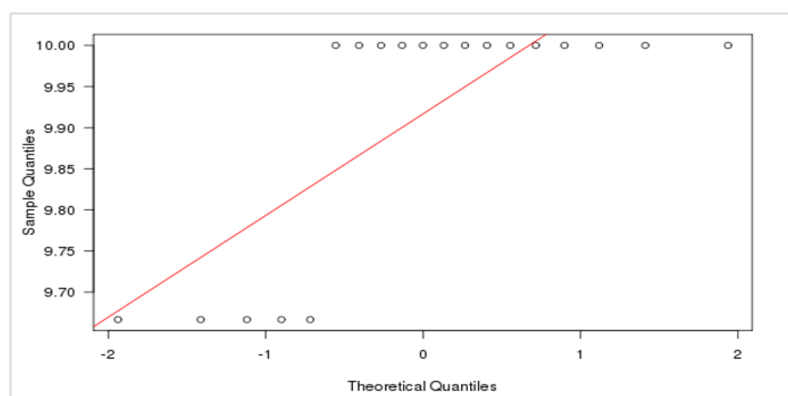


Figure 3. Kruskal-Wallis test.

In detail, figure 3 reveals that the education, linguistics, and social sciences groups expert ratings do not appear to differ statistically significantly, according to a calculated Kruskal-Wallis value of $H = 0.6885$ and a likely non-significant p value ($p > 0.05$). In fact, this suggests that the content validity form was assessed by validators from the three discipline clusters in a generally consistent manner, validating the instrument stability and cross-disciplinary generalizability. Scientifically implying, the outcome supports the claim that any derived

content validity indices represent a common assessment of item quality and relevance across domains rather than being discipline-specific artifacts.

To further establish the normality of data, the computed p-value equals $2.22e-16$ [$P(x \leq 8.1292) = 1$]. It means that the chance of type I error (rejecting a correct H_0) is small: $2.22e-16$ ($2.2e-14\%$). It is therefore assumed that the data is not normally distributed. Since $p\text{-value} < \alpha$, we reject the H_0 . In other words, the difference between the data sample and the normal distribution is big enough to be statistically significant. Similarly, the test statistic W equals 0.2008, which is not in the 95% region of acceptance. Statistically, the observed effect size KS-D is large, 0.5054. This indicates that the magnitude of the difference between the sample distribution and the normal distributions is large.

In the main, figure 4 implies that a p-value of $2.22e-16$ with $P(X \leq 8.1292) = 1$ ($X \leq 8.1292 = 1$) means that if the null hypothesis were true. Meaning that there would be a very little chance of getting the observed test statistic. Thus, this suggests that there is substantial statistical evidence against the null in your study, which would cause it to be rejected at any standard alpha level. In a nutshell, the outcome reinforces the robustness of the content validity form across the evaluated conditions by supporting a real and considerable effect or difference in the aspect being tested (for instance, very high agreement or a strong correlation).

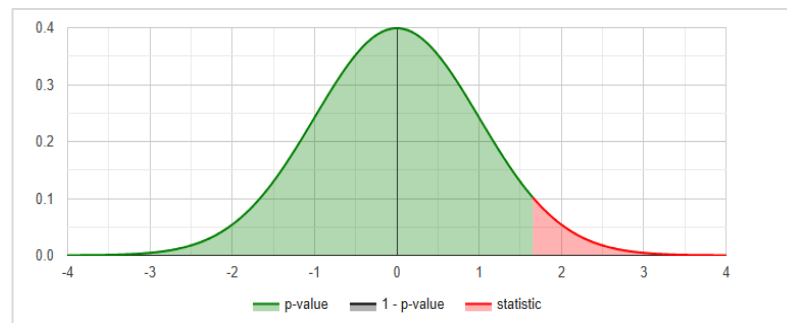


Figure 4. Shapiro-Wilk normality test.

Most essentially, figure 5 discloses the excellent inter-rater agreement, implying a very high agreement among experts on content validity (0.81-1.00 – almost perfect agreement). Additionally, the revealed high agreement reflects the strong consensus on the form items and components, indicating high reliability and consistency of the evaluated form similarly. This likewise confirms that the innovated content validity form has excellent reliability and consistency across experts (Fleiss' kappa = 0.986).

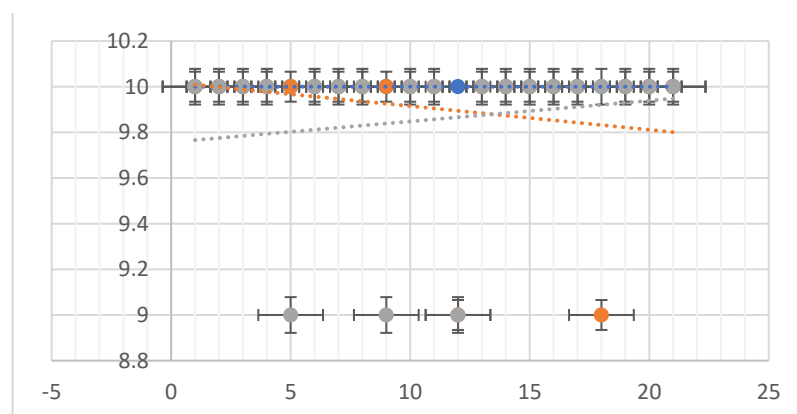


Figure 5. Fleiss' kappa test.

Table 3 indicates the assessment of participants in ensuring the reliability and validity of researcher-made questionnaires prior to its distribution to the target recipients. From the coded responses, it generated five themes such as accuracy, relatability, appropriateness, wide-ranging, and relevance. In fact, the themes that emerged are essential components to be taken into consideration in the fields of education, linguistics, and social sciences.

Table 4. Qualitative assessment of the content validity form.

Verbatim notes/ Transcribed response	Code	Theme
--------------------------------------	------	-------

The CV form has to highlight numerical range to determine the appropriate CVI for accuracy purposes.	To highlight numerical range to determine the appropriate CVI	Accuracy
My overall remark is excellent. This form is very promising in the fields of education, linguistics, social sciences, philosophical & anthropological studies really.	Very promising in the fields of education, linguistics, social sciences, philosophical & anthropo-logical studies	Relatability
I just notice why it started from 7. Well...I think result will be the same, so it is still useful because it has a guide to compute the CVI.	Useful because it has a guide	Appropriateness
I like the qualitative validation part. This will really go an extra mile.	The qualitative validation part.	Wide-ranging
It's good that there is the open-ended question on the last part. This really determine the gray areas for improvement of the questionnaire being validated by experts.	Good that there is the open-ended question on the last part.	Relevance

The findings demonstrate that the developed content validity form possesses strong psychometric properties and is highly applicable across multiple academic disciplines. The instrument achieved excellent content validity ratings, high inter-rater agreement, and consistent evaluations among experts from education, linguistics, and social sciences, indicating its reliability and cross-disciplinary relevance. Although the data exhibited non-normal distribution, the statistical analyses confirmed the robustness and stability of the form. Furthermore, the qualitative feedback reinforced its practical value by highlighting its accuracy, appropriateness, relevance, relatability, and capacity to support comprehensive expert validation. Taken together, the results provide substantial evidence that the content validity form is a reliable and effective tool for evaluating researcher-made questionnaires and strengthening the quality of data collection instruments.

4. Discussion

The outcomes of this validation accentuate the indispensable role of content validity in guaranteeing the value and bearing of research tools across varied academic fields, including instruction, linguistics, social disciplines, philosophy, and similarly, anthropology (Moktar et al., 2025; Rodríguez-Rojas et al., 2024). The documentation and authentication of vital components within these forms disclose both harmonies and discipline-specific distinctions, underscoring the perceived significance of a multidisciplinary method to instrument development. The consistency analysis further validates that well-structured content validity forms will most likely offer reliable and honest results correspondingly, supporting their applicability in research and evaluation settings (Nyarugwe & Jespersen, 2024; Ramezani & Mostafavi, 2025). A fundamental discovery is the documentation of central components that reliably appear across cognitive areas. These comprise lucidity of objectives, bearing of items, inclusiveness of coverage, and connection with theoretical contexts (Anchunda & Kaewurai, 2025; Bingham, 2023). In teaching, for instance, clearness and placement with curriculum ideals are supreme, whereas in linguistics, the emphasis often changes toward the accuracy of language paradigms and the suitability of philological tasks. Social sciences highlight the circumstantial bearing of items, guaranteeing that they really mirror real-world phenomena and diverse standpoints. Meanwhile, in philosophy and anthropology, it prioritizes theoretical depth and social sensitivity, warranting that items are not just valid but likewise reverential to varied epistemological and social upbringings (Anjum & Aziz, 2024; Herzog, 2025).

The consistency analysis undertaken in this inquiry indicates that content validity forms, once accurately designed and authenticated, display high internal reliability and inter-rater consistency. This has been predominantly apparent in the numerical measures, like Cronbach's alpha and intra-class correlation quantities, which show robust agreement amongst professional validators. The qualitative response from field experts supplements these results, with raters reliably noticing the clearness, completeness, and significance of the form. Nevertheless, some inconsistency was detected across various areas, signifying that while the central features are generally significant, their operationalization may vary based on disciplinary models and research objectives (Gierus et al., 2025; Lim, 2025). Moreover, one noteworthy insight from the professional validation procedure is the reputation of iterative fine-tuning. Several validators recommended modifications to advance the lucidity and specificity of some parts, mainly in fields where intangible concepts are predominant adaptations (Hartelt & Martens, 2025; Osborn et al., 2025). For example, in philosophy,

validators may suggest the insertion of more overt descriptions and instances to warrant that items are construed reliably. In anthropology, a validator may also accentuate the necessity for social contextualization, signifying that items must be modified to imitate the exact cultural frameworks under investigation. These suggestions underscore the worth of continuing dialogue and partnership among professionals from diverse disciplines, as well as the necessity for mobility in the plan and application of content validity procedures (Khatri et al. 2024; Maleki et al. 2025).

Most significantly, this paper correspondingly discloses the challenges in realizing harmony across disciplines. Although there is wide-ranging agreement on the significance of central components, variances in disciplinary priorities and epistemological expectations may result in conflicting interpretations of what establishes valid and consistent content (Klein, 2020; Schutte et al., 2025). For instance, in the education area, there is frequently a strong consideration on real-world pertinence and position with learning results; however, in philosophy, the emphasis could be more on theoretical rigor and academic rationality. These perceived variances underline the necessity for a well-adjusted approach that complements disciplinary multiplicity while similarly encouraging cross-disciplinary discourse and partnership (Bräbler, 2025; Leixnering et al., 2026).

The findings suggest that content validity serves not only as a technical measure of instrument quality but also as an indicator of scholarly consensus across disciplines. The high validity ratings, strong inter-rater agreement, and consistent evaluations among experts demonstrate that the proposed form captures universally recognized characteristics of effective research instruments. Moreover, the qualitative themes of accuracy, relevance, appropriateness, and relatability indicate that experts view content validity as essential for ensuring the collection of meaningful, reliable, and contextually appropriate data.

5. Conclusions

The paper demonstrate that the suggested content validity form produces trustworthy and consistent expert assessments in the social sciences, linguistics, and education. While non-significant variances between disciplinary clusters suggested cross-field stability rather than discipline-bound bias, consistently high CVI/CVR values and excellent inter-rater agreement supported the items adequacy, clarity, and representativeness. Hence, the form practicality and interpretive clarity for users were improved through iterative revision made possible by the integration of quantitative indices and qualitative feedback. Collectively, the results point to the content validity form as a reliable instrument for methodically recording and assessing content evidence in a variety of social science and educational situations. Most importantly, the scientific findings contained in this paper, will potentially help researchers in the fields of education, linguistics, and social sciences to ensuring the content validity and reliability of their developed questionnaires.

From the validation component and computation of reliability indexes of the developed content validity form through the online Cronbach's alpha calculator, Kruskal-Wallis, Shapiro-Wilk normality test, and Fleiss' kappa to determine the normality of collected data, this paper confirms that content validity forms are indispensable tools for guaranteeing research excellence across the fields of education, dialectology, social sciences, philosophy, and anthropology. The recognized components, such as lucidity, bearing, inclusiveness, and alignment, are determining factors for worldwide applicability and discipline-specific variations. Reliability examination confirms that well-developed forms produce reliable outcomes, while professional authentication highlights the position of iterative improvement and multidisciplinary partnership. These discoveries offer a robust basis for developing effective and dependable instruments, sup-porting laborious research and evaluation practices in varied academic arenas.

Lastly, the recognized components such as lucidity, bearing, inclusiveness, and alignment, determine both applicability and discipline-particular variations. Reliability examination vouches that well-developed forms produce reliable outcomes, while professional authentication highlights the position of iterative improvement and multidisciplinary partnership. These discoveries offer a robust basis for developing effective and dependable instruments, sup-porting laborious research and evaluation practices in varied academic arenas. Researchers now recommend that the newly developed tool be used within the identified disciplines to maximize its benefits in field research

Supplementary Materials & Data Availability Statement: Future readers are advised to send formal

communication to author 1 if they desire to request for supplementary data.

Author Contributions: Prof Estremera: conceptualization, editing, writing, finalization, and publication. Prof Sevilla and Jeresano: computation of CVI, Cronbach Alpha calculator, Shapiro-Wilk normality test, and Fleiss' kappa test of reliability agreement. Prof Cantoria and Acosta: collection and collation of data

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study in form of written letter sent 2nd week of October 2025.

Acknowledgments: Authors are profoundly thankful to the validators of this academic undertaking which resulted to the successful completion and eventual publication. Authors would also like to acknowledge their affiliate institution for their campaign towards research intensification for professional growth and university's aim to be internationally known through research undertakings.

Conflicts of Interest: The authors declare no conflict of interest in any form possible.

Appendix A. Content validity form.

Content validation criteria	7	8	9	10
<i>I. Quantitative measure</i>				
1. The questionnaire is brief and limited in scope that target participants would feel easy to accomplish without having to require much of their precious time answering.				
2. The questionnaire bears clarity and conciseness of instructions that potential respondents will not exert much effort in trying to grasp them.				
3. The questionnaire is geared towards collection of accurate data and aligned with appropriate research design; otherwise, it will most likely collect flawed and inaccurate data which may heavily affect research outcome.				
4. The questionnaire considers sensitivity and being compassionate to the respondents and does not have humiliating and personal questions which may make target users to be uneasy and apprehensive while answering the questions.				
5. The developed questionnaire bears the element of being time- oriented which implies that it will not take too much effort accomplishing and responding.				
6. The questionnaire is within the ambit and scope of the topic being explored. It will gather appropriate and essential data from the target participants.				
7. The questionnaire promotes flexibility and adaptableness to certain unlooked-for conditions that respondents can still be able to respond.				
8. The developed questionnaire has the element of relatability to the target audience making the included questions easy to answer.				
9. The questionnaire undergoes proofreading for possible lexical and grammatical errors.				
10. The questionnaire considers legibility and does not have too much colors which may distract the target users.				
11. The questionnaire fosters sensitivity to gender orientation, age, ethnicity, and religion.				
12. The questionnaire offers ample options where respondents can comfortably choose the most relevant answer from the given choices.				
13. The statements included in the questionnaire consider the principles of truthfulness and reliability.				
14. The questionnaire could be administered to a wider audience and varied groups to achieve higher validity of research findings.				
15. The questionnaire is well-matched with other ensuing data such as transcriptions and datasets resulting to a substantial raw data capturing completely the research site.				
16. The questionnaire does not demand the intended users to calculate higher statistics just to satisfy what is asked for in the questionnaire.				
17. The questionnaire observes alignment with research objectives, research questions and hypothesis.				

18. The questionnaire does not force the respondents to give personal information thereby respecting the privacy option of the target participants.

19. The questionnaire takes into consideration originality of information which may redound to a better grasp of existing policy, issues and concerns, and/or program.

20. The questionnaire wisely considers statements that will resonate both in local and international perspectives contributing to the body of knowledge.

II. Qualitative notes

1. How do you assess the overall validity and quality of the instrument? What are your further comments for improvement?

2. Please give your honest remarks about the alignment of the questionnaire/ survey form to the research questions, hypotheses/assumptions.

3. How will the questionnaire's data contribute to the existing body of knowledge and international audiences? Please indicate the parts needing refinement.

Note: Revised version as of November 11, 2025.

Appendix B. Content validity tally form.

Contend validation criteria	V1	V2	V3	V4	CVIp
<i>I. Quantitative measure</i>					
1. The questionnaire is brief and limited in scope that target participants would feel easy to accomplish without having to require much of their precious time answering.					
2. The questionnaire bears clarity and conciseness of instructions that potential respondents will not exert much effort in trying to grasp them.					
3. The questionnaire is geared towards collection of accurate data and aligned with appropriate research design; otherwise, it will most likely collect flawed and inaccurate data which may heavily affect research outcome.					
4. The questionnaire considers sensitivity and being compassionate to the respondents and does not have humiliating and personal questions which may make target users to be uneasy and apprehensive while answering the questions.					
5. The developed questionnaire bears the element of being time-oriented which implies that it will not take too much effort accomplishing and responding.					
6. The questionnaire is within the ambit and scope of the topic being explored. It will gather appropriate and essential data from the target participants.					
7. The questionnaire promotes flexibility and adaptableness to certain unlooked-for conditions that respondents can still be able to respond.					
8. The developed questionnaire has the element of relatability to the target audience making the included questions easy to answer.					
9. The questionnaire undergoes proofreading for possible lexical and grammatical errors.					
10. The questionnaire considers legibility and does not have too much colors which may distract the target users.					
11. The questionnaire fosters sensitivity to gender orientation, age, ethnicity, and religion.					
12. The questionnaire offers ample options where respondents can comfortably choose the most relevant answer from the given choices.					
13. The statements included in the questionnaire consider the principles					

of truthfulness and reliability.

14. The questionnaire could be administered to a wider audience and varied groups to achieve higher validity of research findings.

15. The questionnaire is well-matched with other ensuing data such as transcriptions and datasets resulting to a substantial raw data capturing completely the research site.

16. The questionnaire does not demand the intended users to calculate higher statistics just to satisfy what is asked for in the questionnaire.

17. The questionnaire observes alignment with research objectives, research questions and hypothesis.

18. The questionnaire does not force the respondents to give personal information thereby respecting the privacy option of the target participants.

19. The questionnaire takes into consideration originality of information which may redound to a better grasp of existing policy, issues and concerns, and/or program.

20. The questionnaire wisely considers statements that will resonate both in local and international perspectives contributing to the body of knowledge.

Σ/N

Overall CVI- Σ/N

Legend: V1 – expert 1; V2 – expert 2; V3 – expert 3; V4 – expert 4; V5v – Expert 5; CVIp – content validity index per component; CVI – content validity index; Σ/N – summation divided by the number of cases.

Notes:

- (i) Please seek permission from the author 2 if you are to adopt this innovative form. Proper acknowledgement and citation must be made. Please send a communication to michaelestremera@debesmscat.edu.ph
- (ii) You may add as many columns as possible depending on the number of expert validators.
- (iii) There are 2 ways of getting the overall CVI. Either must have same value.
- (iv) Likert Scale: (7 – Not valid; 8 – Fairly valid; 9 – Moderately valid; 10 – Highly valid).

References

- Anchunda, H. Y., & Kaewurai, W. (2025). An instructional model development based on inquiry-based and problem-based approaches to enhance prospective teachers' teamwork and collaborative problem-solving competence. *Social Sciences & Humanities Open*, 11, 101480. <https://doi.org/10.1016/j.ssaho.2025.101480>
- Anjum, G., & Aziz, M. (2024). Advancing equity in cross-cultural psychology: Embracing diverse epistemologies and fostering collaborative practices. *Frontiers in Psychology*, 15, 1368663. <https://doi.org/10.3389/fpsyg.2024.1368663>
- Aslan, M., Sağtaş, S., & Özkaynar, K. (2025). Metaverse Perception Scale Development and Validation. *Sage Open*, 15(2). <https://doi.org/10.1177/21582440251329126>
- Bingham, A. J. (2023). From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231183620>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6, 366616. <https://doi.org/10.3389/fpubh.2018.00149>
- Braßler, M. (2025). Interdisciplinary learning for corporate social responsibility and sustainability across business and psychology. *Discover Sustainability*, 6, 1109. <https://doi.org/10.1007/s43621-025-02078-9>
- Carmines, E. G., & Zeller, R. A. (1979). Validity. In *Validity* (pp. 17-27). SAGE Publications, Inc., <https://doi.org/10.4135/9781412985642.n2>
- Dalawi, I., Isa, M. R., Chen, X. W., Azhar, Z. I., & Aimran, N. (2023). Development of the Malay Language of understanding, attitude, practice and health literacy questionnaire on COVID-19 (MUAPHQ C-19): content validity & face validity analysis. *BMC public health*, 23(1), 1131. <https://doi.org/10.1186/s12889-023-16044-5>
- Elangovan, N., & Sundaravel, E. (2020). Method of preparing a document for survey instrument validation by experts. *MethodsX*, 8, 101326. <https://doi.org/10.1016/j.mex.2021.101326>
- Gierus, B., Du, T., Maduforo, A. N., Gilbert, B., & Koh, K. (2025). Prevalence and Quality of Mixed Methods Research in Educational Subdisciplines: A Systematic Review. *SAGE Open*. <https://doi.org/10.1177/21582440251335171>
- Govindasamy, P., Cumming, T. M., & Abdullah, N. (2024). Validity and reliability of a needs analysis questionnaire for the development of a creativity module. *Journal of Research in Special Educational Needs*, 24(3), 637-652. <https://doi.org/10.1111/1471-3802.12659>
- Gupta, S., Grover, S., Menon, V., Indu, P. V., Vidhukumar, K., & Chacko, D. (2025). Item generation and establishing face and content validity of a rating scale: A primer. *Indian Journal of Psychiatry*, 67(8), 816-822. https://doi.org/10.4103/indianjpsychiatry_750_25
- Hartelt, T., & Martens, H. (2025). Development and validation of an instrument measuring students' self-efficacy regarding explaining evolutionary changes. *Journal of Biological Education*, 1-35. <https://doi.org/10.1080/00219266.2025.2546795>
- Herzog, C. (2025). Relational & culture-sensitive AI innovation. *Journal of Responsible Technology*, 23, 100135. <https://doi.org/10.1016/j.jrt.2025.100135>

- Khatri, P., Shukla, S., Thomas, A., Shiva, A., & Behl, A. (2024). Towards work life fulfilment: Scale development and validation. *Journal of Business Research*, 186, 115006. <https://doi.org/10.1016/j.jbusres.2024.115006>
- Klein, J. T. (2020). Sustainability and Collaboration: Crossdisciplinary and Cross-Sector Horizons. *Sustainability*, 12(4). <https://doi.org/10.3390/su12041515>
- Leixnering, S., Jancsary, D. C., Ayrault, J., Gehman, J., Gray, B., Linnenluecke, M. K., Meyer, R. E., & Pavez, I. (2026). Collaboration for Tackling Grand Challenges: A Curated Conversation. *Journal of Management Inquiry*. <https://doi.org/10.1177/10564926251330828>
- Lim, W. M. (2025). Theory and theory development: Guidelines for establishing theoretical gaps, foundations, contributions, and implications. *Journal of Business Research*, 202, 115745. <https://doi.org/10.1016/j.jbusres.2025.115745>
- Maleki, A., Khodabakhshzadeh, H., Zeraatpishe, M., & Baghaei, P. (2025). Development and validation of the online formative assessment validity scale (OFAVS) in EFL contexts. *Discover Sustainability*, 6, 471. <https://doi.org/10.1007/s43621-025-01336-0>
- Masuwai, A., Zulkifli, H., & Hamzah, M. I. (2024). Evaluation of content validity and face validity of secondary school Islamic education teacher self-assessment instrument. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2308410>
- Moktar, A., Alias, B. S., & Mohamad Arsad, N. (2025). Leading Change: Validating an Instrument to Measure Action Research Culture and Teacher Quality in Malaysian Schools. *Pegem Journal of Education and Instruction*, 15(4), 32-44. <https://www.pegegog.net/index.php/pegegog/article/view/4159>
- Mollman, S., Muckle, T. J., & Martinez, M. (2024). Intentional Learning Theory: Validating Its Measurement. *SAGE Open*, 14(1). <https://doi.org/10.1177/21582440241232745>
- Nyarugwe, S. P., & Jespersen, L. (2024). Assessing reliability and validity of food safety culture assessment tools. *Heliyon*, 10(12), e32226. <https://doi.org/10.1016/j.heliyon.2024.e32226>
- Osborn, L. J., Garvey, L., Adama, E., & Bhowmik, J. (2025). Development and validation of the Competence Assessment Insights Survey for Australian Registered Nurses (CAISA-RN). *Collegian*, 32(6), 436-448. <https://doi.org/10.1016/j.colegn.2025.09.006>
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content Validity – Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1 – Eliciting Concepts for a New PRO Instrument. *Value in Health*, 14(8), 967-977. <https://doi.org/10.1016/j.jval.2011.06.014>
- Plengkham, B., Rattanasak, S., & Sukserm, P. (2025). 6Ps Guidelines for Constructing an English Questionnaire for a Survey in Social Science Research. *Journal of Education and Learning*, 14(3), 212-220. <https://doi.org/10.5539/jel.v14n3p212>
- Ramezani, S. G., & Mostafavi, Z. S. (2025). Developing and validating a comprehensive scale for accreditation standards and quality assurance in e-learning institutions. *Education and Information Technologies*, 30, 21139-21187. <https://doi.org/10.1007/s10639-025-13587-5>
- Rodríguez-Rojas, M. P., Ramírez-Segura, A. M., Valenzuela-Mora, P., Dampuré, J., & Munévar, F. R. (2024). Critical Thinking Evaluation Scale: Design and Validation in a Colombian Population. *SAGE Open*, 14(4). <https://doi.org/10.1177/21582440241297418>
- Schutte, B. G., Bayram, D., Vennix, J., & van der Veen, J. (2025). Exploring the implementation of challenge-based learning for sustainability education in Dutch secondary education: teachers' experiences. *Environmental Education Research*, 31(6), 1166-1192. <https://doi.org/10.1080/13504622.2025.2458723>
- Shrotryia, V. K., & Dhanda, U. (2019). Content Validity of Assessment Instrument for Employee Engagement. *Sage Open*, 9(1). <https://doi.org/10.1177/2158244018821751>
- Tay, L., & Jebb, A. T. (2018). Establishing Construct Continua in Construct Validation: The Process of Continuum Specification. *Advances in Methods and Practices in Psychological Science*, 1(3), 375-388. <https://doi.org/10.1177/2515245918775707>
- Uddin, M. M. (2025). A philosophical inquiry into knowledge and originality to investigate the prevailing criticism of ChatGPT et al. *Discover Education*, 4, 414. <https://doi.org/10.1007/s44217-025-00805-y>