Design and Validation of Questionnaire to Assess the Validity of Designed Models in the Humanities Using a Mixed Method

Behzad Shoghi[1] , Mahsa Karimi[2]*

Received Date:11/11/2025 *Accepted Date:1/1/2026*

Pp: 71-95

Abstract

The main objective of the present study was to design and validate a questionnaire for evaluating the validity of designed models in humanities using a mixed-method approach. This research is applied in terms of its goal and exploratory mixed-method (qualitative-quantitative) in terms of data type, consisting of qualitative meta-synthesis and quantitative descriptive-analytical (correlation) components. The qualitative study population included all theoretical foundations and relevant literature in domestic and international databases, which were selected using purposive non-random sampling and systematic elimination based on the PRISMA model flowchart. A total of 18 scientific works were chosen. The quantitative study population included all professors and faculty members of humanities departments at universities in Tehran. Based on sample size calculations for confirmatory factor analysis and multi-stage cluster random sampling, 370 respondents were selected as the sample size. Data collection tools included systematic literature review for the qualitative section and a researcher-developed questionnaire with 34 items derived from the qualitative section for the quantitative part. It is worth noting that validity in the qualitative section was calculated using the 27-item checklist based on the PRISMA model, and reliability was calculated using Cohen's kappa coefficient. Additionally, in the quantitative section, validity was assessed through face validity, content validity, and construct validity, while reliability was calculated using Cronbach's alpha and composite reliability. Validity and reliability were confirmed in both qualitative and quantitative sections. Data analysis methods included thematic analysis for the qualitative section and descriptive and inferential statistics (confirmatory factor analysis) using PLS software version 3 and SPSS software version 23 for the quantitative section. The findings indicated that the 34-item questionnaire could be used to assess the validity of models from the perspective of experts. The questionnaire evaluates external validity through components such as purpose (4 items), research method design (4 items), controlling confounding variables (8 items), and alignment (8 items), as well as internal validity through components such as logical review (3 items), expert feedback (4 items), and sensitivity analysis (3 items).

Key Words: Model Validation, Internal Validity, External Validity, Generalizability

[1]- Educational Administration Department, SR.C., Islamic Azad University, Tehran, Iran

[2] - Department of Educational Sciences, Payame Noor University (PNU), P.O. Box19395-4697, Tehran, Iran

* Corresponding Author: Mahsa Karimi Email: mkarimi_ed@yahoo.com

Introduction

In today's era of remarkable scientific and technological transformations, the humanities play an indispensable role as a fundamental pillar in understanding human, cultural, and social relations, thereby contributing significantly to the path of sustainable development. Among these, the design of scientific and practical models within the humanities has increasingly attracted attention, as it aims to provide practical solutions to complex issues. What particularly highlights the importance of this field is the scientific validity of the developed models, which not only ensures the quality of research but also leads to the production of reliable and applicable knowledge (Lim, 2024). In Iran, considering the country's unique social and cultural context and the necessity of localizing such models, the scientific validation of these frameworks holds even greater significance. Without achieving valid and reliable models, the practical utilization of research findings becomes considerably limited.

One of the fundamental challenges in validating designed models within the humanities is the lack of comprehensive and applicable standards for the scientific evaluation of such models. This deficiency has led many researchers to rely on unsystematic or ineffective methods for assessing the validity of their models. The second challenge lies in the lack of transparency in validation criteria; many models are evaluated solely based on the personal judgment of the researcher or the research team, which undermines the reliability of research outcomes. The third challenge stems from the inherent complexity of the humanities, where the multifaceted nature of human interactions and issues makes the validation of models more difficult compared to other fields of science. The fourth challenge is the insufficient attention paid to mixed or integrated validation approaches, which could, by combining the strengths of quantitative and qualitative methods, lead to more precise and comprehensive validation processes. The significance of each of these challenges in humanities research lies in the fact that, without addressing them, the designed models will lack both scientific and practical credibility and, consequently, will be unable to provide effective solutions to societal issues (Bernhard-Harrer et al., 2025).

As noted earlier, the scientific validation of designed models in the humanities not only contributes to enhancing the quality of research but also increases confidence in the results and practical applications of these models within society. By employing valid and standardized validation methods, researchers can develop models that analyze existing issues more accurately and comprehensively while providing feasible and actionable solutions. Moreover, proper model validation allows for the generalization of findings to other societies and similar contexts, ultimately enriching scientific research and fostering the advancement of knowledge in the humanities.

Unfortunately, in many humanities studies, the validation of designed models is conducted in an unscientific and ineffective manner. This has resulted in the development of models that are not only unreliable but also incapable of being effectively applied to solve real-world problems. Issues such as the use of non-standard validation methods, the absence of clear evaluation criteria, and the lack of attention to the inherent complexities of the humanities have caused many existing models to remain purely theoretical and detached from practical application. This unfavorable situation has motivated further research in this area, aiming to develop scientifically grounded validation methods that can lead to the creation of credible and practically applicable models within the humanities.

The failure to address issues related to the validation of designed models in the humanities entails numerous negative consequences. First, the lack of confidence in research findings and the resulting decline in their scientific credibility undermine both academic and public trust in humanities research. Second, invalid models are incapable of effectively addressing real-world problems or providing practical solutions, leading to the waste of valuable resources and time. Third, the absence of valid models leaves many social and cultural issues unresolved, thereby exacerbating existing challenges. Fourth, neglecting the scientific validation of models hinders the progress and development of the humanities as one of the fundamental pillars of sustainable development (William, 2024).

The present study can have far-reaching impacts on addressing existing challenges and advancing knowledge within the humanities. First, by introducing rigorous and scientific methods for validating designed models, this research contributes to improving the overall quality of studies in the humanities. Many existing models in this field lack sufficient reliability due to the absence of standardized validation tools. Through the development of a valid and reliable questionnaire, this study enables more precise model evaluation and provides researchers with a scientific and trustworthy instrument to assess the quality and validity of their models.

The second contribution of this research lies in enhancing the scientific understanding of the model design process in the humanities. Due to the social and cultural complexities inherent in this domain, model development requires comprehensive and precise approaches. The design and validation of a questionnaire capable of evaluating the credibility of designed models help researchers engage in model development with greater confidence and awareness of the strengths and weaknesses of their proposed frameworks. This, in turn, ensures that designed models are not only theoretically sound but also practically applicable.

Third, by presenting a scientific framework for model validation, this study contributes to the expansion of knowledge in the humanities. The mixed-methods approach employed herein—integrating both quantitative and qualitative methodologies—allows for more comprehensive and accurate analysis. This combination not only enhances the quality of research findings but also advances methodological development in the humanities and promotes the broader adoption of integrated approaches in this field.

Fourth, conducting this study has the potential to bring about positive transformations within the humanities. By providing validated models and reliable validation instruments, researchers will be able to study social, cultural, and human issues with greater precision and propose solutions that are both feasible and impactful. Furthermore, the findings of this research can improve evidence-based decision-making processes in related fields and enhance the practical applications of humanities models in policymaking, education, and other domains.

Overall, beyond its direct impact on the quality of humanities research, this study can pave the way for broader transformations in methodology, model design, and the practical application of humanities knowledge in society. Ultimately, the unresolved issues surrounding model validation not only undermine the quality of research and designed models but also significantly impede the development and practical utilization of the humanities. Therefore, addressing the scientific validation of designed models is a vital necessity for improving the quality and effectiveness of humanities research.

Theoretical Foundations

A model is designed to describe and analyze systems and to provide theoretical predictions, whereas a pattern focuses on identifying and representing variations or relationships within data. The choice between using a model or a pattern depends on the research objectives, the type of data, and the desired outcomes of the study. In other words, if the goal is to achieve an analytical and theoretical understanding of a phenomenon, a model is more appropriate. However, if the aim is to identify trends or behaviors within datasets, a pattern would be the preferred choice (Varipuu et al., 2020).

A model and a pattern are two distinct concepts used in research, each with its own unique characteristics and applications (Fred, 2020). The following section outlines their differences and the contexts in which each is used:

Definition, Characteristics, and Application of Models and Patterns

A model is a simplified or conceptual representation of a system, theory, or process, designed to explain, analyze, or predict the behavior of systems. Models often consist of variables and the relationships among them. They can be developed in either quantitative form (e.g., mathematical models) or qualitative form (e.g., conceptual models). The main purpose of a model is to predict, analyze, or represent complex relationships.

The application of a model is appropriate when there is a need to describe and predict the behavior of a system or phenomenon, to test theories and hypotheses, or to simulate various outcomes based on changes in variables. Models are particularly useful in complex analyses involving interactions among multiple variables.

A pattern, on the other hand, refers to a recognizable structure or trend within data or phenomena that indicates a specific repetition or relationship. Patterns may be observed either qualitatively or quantitatively and generally involve identifying recurring forms or regularities within datasets. They are typically derived from empirical evidence and observations and can contribute to understanding historical behaviors or forecasting future developments. Patterns are often associated with the identification of trends and relationships in collected data. It is important to note that patterns are primarily utilized when the research objective is to identify specific trends, behaviors, or relationships within data.

When the goal is to examine common methods or specific behaviors within a dataset, or to analyze data for detecting behavioral patterns or particular distributions, the concept of a pattern becomes essential.

Model Validity

Model validity refers to the model's ability to accurately predict or describe real-world phenomena. Assessing and verifying model validity is a multi-stage process that can be carried out through various methods. Evaluating model validity helps identify the model's strengths and weaknesses, ensuring its accuracy and consistency with real-world conditions and existing theories.

Model validity is generally categorized into two main types: internal validity and external validity, both of which play a crucial role in assessing the quality of the outcomes produced by designed models. The following sections explain each type of validity and the methods used to evaluate them (Shanbaum, 2024).

1. Internal Validity of the Model

Internal validity refers to the accuracy and robustness of the relationships within the model. In other words, internal validity examines whether the model—or the conducted research—correctly represents the relationships among the variables and whether the obtained results are a true reflection of actual relationships rather than being influenced by irrelevant or random factors.

The internal validity of a model can be assessed using the following methods:

1. Hypothesis Testing of the Model:

Model validity can be evaluated by testing the hypotheses regarding the relationships among variables. In this process, the findings derived from thematic analysis, which led to the model's design, are examined using statistical tests such as Confirmatory Factor Analysis (CFA) and Structural Equation Modeling (SEM). This step ensures that the data collected from interviews or observations align with the theoretical assumptions underlying the model, and it verifies whether the results indicate significant and meaningful relationships among the model's components.

2. Additional Methods for Assessing Internal Validity

1. Use of Various Charts:

Different charts representing the data and models derived from analyses are created to examine whether they accurately reflect the findings that led to the model's development. In other words, these visualizations help determine whether the charts and data representations confirm the relationships and structure of the model's components.

2. Logical Review:

Evaluating the model from a logical perspective—ensuring its consistency with established theories and scientific principles—contributes to internal validity. The model should appear logically coherent in terms of the relationships among its components and be compatible with existing theoretical foundations.

3. Expert Feedback:

Obtaining experience-based feedback from aligned experts and researchers can aid in refining and improving the model and validating its assumptions. This involves reviewing whether expert opinions strengthen or suggest modifications to the model and its relationships. Techniques such as the Delphi method can be employed to systematically gather and integrate expert insights into the validation process.

4. Sensitivity Analysis

Sensitivity analysis examines whether small changes in the model's inputs—such as thematic analyses of conducted interviews or expert opinions—affect the final results and the relationships among the model's components. Essentially, this analysis assesses whether the model's outcomes are stable or highly sensitive to variations in its inputs.

The primary goal of sensitivity analysis is to determine the extent to which minor changes in the model's inputs can influence the final results and the interrelationships between components. In simpler terms, small changes in model inputs may include:

Examples of Small Changes in Sensitivity Analysis

1. Changes in Input Data

Example: If expert opinions are used as input data, altering one or more opinions can significantly impact the results. For instance, if a single expert changes their view on a component, does this adjustment affect the final outcome?

2. Changes in Weights or Importance of Components

Example: Some factors or components may be assigned greater weight. If the weight of a component is slightly adjusted (e.g., from 0.3 to 0.4), does this change influence the overall results?

3. Changes in Assumptions

Example: Suppose a specific relationship between two components is assumed. If this assumption is modified slightly (e.g., from a "positive relationship" to a "negative relationship"), could this change significantly alter the model's outcomes?

4. Changes in Data Collection Methods

Example: Using different methods to gather data (e.g., interviews, questionnaires, or observation) may affect results. For instance, if shallow questionnaires are used instead of in-depth interviews, would the results differ substantially?

5. Changes in Timing or Conditions of Data Collection

Example: Data collected at different times or under varying conditions (e.g., before or after a particular event) may affect results. Would such variations influence the final outcomes?

Summary:

In sensitivity analysis, the goal is to examine how small changes in the model's inputs—such as expert opinions, component weights, or assumptions—can affect the final results. This process provides a clearer understanding of the model's stability and overall validity.

2.External Validity of the Model

External validity refers to the ability to generalize the results obtained from a study to other populations, times, locations, or conditions. In other words, it indicates whether the findings of a research study hold true in different contexts or with different samples (Classa et al., 2024; Larinzi et al., 2024). It is important to note that when designing a model—especially in social, economic, and cultural contexts—considering the model's external validity is of critical importance.

Methods for Assessing External Validity of the Model

1. Reproducibility:

The ability to replicate research results under similar conditions using similar methods. If a study's results are reproduced across different studies and with different samples, its external validity is strengthened. Various model fit indices are used for this purpose, which indicate whether the model's findings are consistent with real-world data.

2. Predictive Power Testing:

Assessing the model's ability to predict outcomes in new or different datasets (new conditions) contributes to evaluating its external validity. This method examines the model's accuracy in forecasting future results by comparing the designed model with real-world data, answering the question of whether the model can reliably predict new facts and data.

3. Comparison with Previous Empirical Studies:

Evaluating whether the obtained results align with findings from prior research. The designed model is assessed against relevant research documentation to determine if previous studies support the model. This method addresses whether the model remains stable and consistent across different times and contexts.

4. Comparison with Previous Theoretical Studies:

Examining whether the results are consistent with prior theoretical findings. The model is evaluated against relevant theoretical documentation to verify whether existing theoretical foundations support it. This approach also assesses the model's stability and consistency over time.

5. Comparison with Other Models:

The proposed model is compared with existing models to analyze its strengths and weaknesses. This method addresses whether the new model produces better results than other models and whether its findings are compatible with other models, thereby supporting generalizability to different contexts.

6. Control of Confounding Variables for Generalizability:

Implementing the model under different environmental conditions, locations, and times and obtaining consistent results indicates external validity. The model should be designed to identify and control confounding variables—such as geographic conditions, cultural, social, political, economic, legal, and religious factors—that may influence outcomes, ensuring it can be generalized to other contexts.

7. Designing an Appropriate Research Method:

When designing a suitable research method, the following should be considered:

• Selection of an appropriate type of research for model design: Choosing a research type based on paradigms, data type, objectives, and data collection environment helps achieve better understanding of the results and facilitates their generalization to other populations.

- Use of Appropriate Sampling Methods and Sample Size in Model Design:

The sample should be selected to represent the entire population, incorporating diverse characteristics. If the sample accurately reflects the target population, the model's findings are more likely to be generalizable. In other words, diversity within the sample enhances the ability to generalize results to different groups and increases the model's external validity.

- Use of Appropriate Measurement Instruments in Model Design:

Employing valid and reliable instruments for data collection contributes to external validity. Valid and reliable tools ensure that the data are accurately collected, thereby enhancing confidence that the results can be generalized to other contexts.

- Use of Appropriate Statistical Methods for Data Analysis in Model Design:

Applying suitable statistical techniques improves the understanding of research results and facilitates their generalization to other populations. Proper statistical analysis ensures that findings are robust and can be applied beyond the original study sample.

Methodology

Research Type

In this study, the type of research is defined as follows:

- Based on Objective: Applied research.
- Based on Data Type: Sequential mixed-methods with an exploratory approach (qualitative–quantitative).
- Based on Paradigm: Pragmatic or combined (integrating interpretivism and positivism).
- Based on Nature (Approach and Design): In the qualitative phase, an integrative approach was used, followed by a descriptive-analytical approach in the quantitative phase.
- Based on Reasoning Type (Logic of Execution): Mixed-methods research (inductive–deductive), as it employs both inductive reasoning (in the qualitative content analysis phase) and deductive reasoning (in the quantitative correlation phase).

Population and Sample (Size and Sampling Method)

a. Qualitative Phase:

The population for the qualitative phase, using the integrative review approach, included all articles and scholarly works available in domestic and international databases, as well as relevant documents and regulations in the field. In this phase, 18 articles were selected through purposeful non-random sampling based on the article selection process outlined in the PRISMA guidelines.

It is noteworthy that the inclusion criteria for articles in the integrative review method comprised:

- Relevance to the research topic and recency of publication,
- High scientific quality and credibility, sourced from reputable domestic and international databases,
- Appropriate research methodology,

- Diversity of perspectives and viewpoints, among other considerations.

b. Quantitative Phase:

In the quantitative phase, the population consisted of faculty members and academic staff in the humanities disciplines at universities in Tehran. Considering that prominent scholars, including Klein (2015), recommend a minimum sample size of 200 participants as a general rule for Confirmatory Factor Analysis (CFA), this study, aiming to conduct CFA and enhance generalizability, selected 370 respondents using multi-stage cluster sampling.

The questionnaire was distributed online, and 13 incomplete responses were excluded from the analysis. Consequently, statistical analyses were performed on 357 valid responses.

Data Collection Tools, Validity, and Reliability

a. Qualitative Phase:

In the first phase of the study, using the integrative review approach, data collection was conducted through a systematic review of the literature and reputable scientific sources. This process involved a precise and targeted search of scientific databases, articles, books, and theses relevant to the research topic.

Content Validity:

The content validity of the integrative review was confirmed, demonstrating that the concepts and content under investigation were comprehensively covered in the existing literature. Articles were carefully selected, beginning with a thorough screening process. A flow diagram was used to track the search and selection of appropriate articles for the study, considering restrictions such as:

- Temporal scope: domestic and international publications,
- Geographic scope: domestic and international databases,
- Research nature: synthesis, review, and qualitative studies,
- Subject scope: keywords used for searching.

The screening process included both coarse and fine filtering stages. Additionally, internal validity analysis showed that the findings of the integrative review were not influenced by external factors and were accurately interpreted. The process involved:

- A 27-item checklist based on the PRISMA model,
- Independent analysis by the researcher and a statistics expert,
- Cohen’s Kappa coefficient to assess inter-rater agreement,
- Use of standard criteria and reproducibility (transparency in method execution),
- Utilization of MAXQDA software for precise tracking of analysis and coding steps,
- Review, feedback, and revision of codes by a subject-matter expert to identify inconsistencies.

Reliability:

Reliability focused on the stability and repeatability of results. Articles were independently analyzed by the researcher and a statistics expert, with Cohen’s Kappa indicating a high degree of agreement. Furthermore, the 27-item PRISMA checklist was applied as a standard criterion for assessing the quality and reliability of data, covering sections such as abstract, methods, findings, and discussion, enhancing transparency and precision in reporting.

Finally, by clearly documenting the stages of article selection, analysis, and data extraction according to the PRISMA protocol, the study ensured that its procedures could be replicated and reviewed by other researchers. This demonstrates that the systematic review process was designed to achieve high content and internal validity, as well as reliability and reproducibility.

b. Quantitative Phase:

To assess the validity of the questionnaire, both content validity and construct validity were evaluated.

Content Validity:

Using Lawshe's method, the Content Validity Ratio (CVR) and Content Validity Index (CVI) were assessed with the help of 20 research experts. The questionnaire content was reviewed to identify redundant questions and to suggest modifications where necessary. During this process:

- 12 items were revised based on expert feedback to improve simplicity, clarity, and relevance,
- 4 items were deemed unnecessary by the experts and removed from the questionnaire.

Construct Validity:

Construct validity was assessed through convergent and discriminant validity using SmartPLS 3 software. The results confirmed the construct validity of the questionnaire, as summarized in the corresponding table.

Reliability:

Reliability was evaluated using Cronbach's alpha, composite reliability, and McDonald's omega. The reliability coefficients for all questionnaire constructs exceeded 0.7 (Hair et al., 2017), indicating that the measurement instrument was reliable. It is noteworthy that in studies employing Structural Equation Modeling (SEM), McDonald's omega is preferred due to its higher accuracy in assessing reliability (McDonald, 1999).

The validity and reliability coefficients for the questionnaire are presented in Table 1.

Table 1 – Distribution of Questionnaire Indicators and Their Validity and Reliability (Source: Researcher-Developed Instrument)

Component	α	CR	ω	AVE	MSV	ASV	HTMT
Objective	0.73	0.80	0.83	0.56	0.41	0.23	0.68
Research Method Design	0.78	0.82	0.85	0.58	0.43	0.25	0.64
Control of Confounding Variables	0.72	0.81	0.82	0.64	0.46	0.30	0.73
Alignment	0.75	0.79	0.84	0.61	0.45	0.29	0.71
Logical Review	0.71	0.83	0.81	0.63	0.42	0.22	0.74
Expert Feedback	0.74	0.86	0.80	0.67	0.47	0.26	0.70
Sensitivity Analysis	0.77	0.84	0.86	0.59	0.44	0.31	0.69

The findings in the table indicate that the AVE values for all seven components are significantly above 0.5, suggesting that more than half of the variance of each construct is explained by its corresponding items, thereby supporting convergent validity.

Regarding discriminant validity, the Fornell-Larcker criterion confirms that the AVE of each construct is greater than the squared correlations with other constructs. This indicates that each dimension is well-distinguished from the others, showing no overlap. Additionally, the Maximum Shared Variance (MSV) and Average Shared Variance (ASV) for all constructs are substantially lower than the respective AVE values. This further demonstrates that each construct independently and effectively measures its intended concept, confirming discriminant validity.

Specifically:

- When the MSV of a construct is lower than its AVE, the maximum shared variance between that construct and others is meaningfully less than its own variance, indicating that each dimension effectively captures its unique characteristics with minimal overlap.
- ASV values being lower than the AVE indicate that the average shared variance between each construct and the others is also low, enhancing the accuracy and reliability of the measurement instrument.

Moreover, cross-loadings were examined as another indicator of discriminant validity. Results showed that each item loads highly on its respective construct and exhibits minimal cross-loading on other constructs, confirming that there is no interference between components and that each construct is measured accurately and independently.

It should be noted that in the studies conducted by Bandani Pour (2024), Hossein Pour (2024), Hosseinzadeh (2024), Ansarian (2025), Darabian (2025), Ghasemian (2025), Shadalooi (2025), and Parafkandeh Haghighi (2025), this questionnaire was used to assess model validity, and its validity and reliability were calculated. The findings from these studies indicated that the questionnaire is both valid and reliable.

Data Analysis Method

Qualitative Phase:

To identify the dimensions, components, and indicators related to the validity of designed models in the humanities, a mixed-method approach was employed using flexible thematic analysis (Braun & Clarke, 2020) with MAXQDA Analytics Pro 2018 software. Themes and common patterns were extracted from the texts of selected articles.

In the first integrative review phase, based on the findings from this method, the dimensions, components, and indicators related to the validity of designed models in the humanities were identified. Using a systematic review approach guided by the PRISMA model, the integrative review facilitated the identification of these dimensions. The process consisted of the following steps:

1. Steps for Conducting the Research Synthesis to Identify Components Related to Model Validity in Humanities (Mixed-Method):

Phase 1: Defining the Research Geography and Identifying Studies for Use

a) Determining search parameters such as publication date and research type
b) Setting inclusion criteria for collected documents from the previous step
c) Developing search strategies and selecting databases
Phase 2: Systematic Critique of Selected Documents
a) Coarse screening
b) Fine screening
c) Detailed analysis
Phase 3: Synthesis – Creating Something New from Separate Elements
Two types of synthesis were conducted in this phase:
1. Aggregative Synthesis:
This is conceptually opposed to integrative research synthesis (Gough et al., 2012). Aggregative synthesis is analogous to a physical change, where the findings of selected studies are combined, similar to meta-analysis of quantitative research.
2. Integrative (or Transformative) Synthesis:
In this approach, the findings of other studies themselves become data that are combined with other data and then recreated into a new entity. This process allows for the development of new insights and a reinterpreted understanding of the research topic.
2. Flow Diagram (Article Search Process):
In this phase, the following limitations and parameters were first established for the article search:
• Temporal scope: Domestic publications from 2017 to 2025 (1396–1404) and international publications from 2016 to 2025
• Geographic scope: Domestic and international databases
• Research nature: Synthesis, review, and qualitative studies
• Subject scope: Keywords relevant to the research topic
After applying these limits, a coarse and fine screening process was performed. Following the PRISMA flow diagram, a total of 18 articles were selected, and their quality was reviewed and analyzed.
3. 27-Item Checklist for Assessing Article Quality:
The results of the comprehensive search of relevant journals revealed that, between 2017–2025 (2016–2025 internationally), only 30 articles—both using integrative and non-integrative approaches—were published in this domain. The overall conformity of article quality with the checklist criteria was estimated at 64%. The largest quality deficiencies were observed in the methods section, accounting for 54%. Key shortcomings in systematic review reporting included errors in primary studies, errors arising from combining results, and lack of discussion regarding potential biases.
Most identified articles were published between 2017–2020 (64.7%) and 2017–2020 for domestic journals (68.2%), with 46.5% in psychometric studies and 34.1% in educational research. Of these, 33.3% were domestic and 66.6% were international.
The 27-item checklist results indicated that all selected articles were of either high or adequate quality, as quality scores per item were either above 75% or between 50–75%. Items scoring below 50% were considered low-quality.

Inter-Rater Reliability:
Cohen's Kappa was calculated to assess agreement between two evaluators, each classifying N items into C mutually exclusive categories. The obtained Kappa value (0.59) indicates a satisfactory level of agreement between the two evaluators.
4. Analysis and Synthesis (Aggregative and Integrative):
Finally, synthesis was conducted to identify the dimensions, components, and indicators related to the validity of designed models in the humanities using a mixed-method approach. Based on the integrative review and subsequent analyses, the extracted indicators for model validity were obtained.
Quantitative Phase:
In the quantitative phase, given the research objectives, descriptive and inferential statistical methods were employed.

- Descriptive Statistics: Used to describe the demographic characteristics of the participants, including age, gender, education, and work experience, using frequencies, tables, and charts. For the research variables, measures such as mean, standard deviation, skewness, and kurtosis were calculated.
- Inferential Statistics: Confirmatory Factor Analysis (CFA) was conducted to test the construct validity of the questionnaire and the proposed model.

These analyses were performed using IBM SPSS Statistics 23 (2015) and SmartPLS 3 (2016) software.

Research Findings

In this section, the demographic characteristics of the participants and respondents are presented in the following table:
Table 2 – Demographic Characteristics of Participants in the Qualitative and Quantitative Phases

Demographic Variable	Frequency	Percentage
Gender		
Female	175	49%
Male	182	51%
Age Range		
30–40 years	189	53%
40–50 years	96	27%
Above 50 years	72	20%
Work Experience		
Less than 10 years	118	33%
More than 10 years	224	67%

The table above shows the demographic characteristics of the participants in terms of gender, age range, and work experience.
To identify the components of model validity in designed models in the humanities using a mixed-method approach, as previously mentioned, flexible thematic analysis (Braun & Clarke, 2020) was applied. In this stage, key themes and concepts in the data were identified and labeled.
The researcher focused on identifying and labeling phrases, keywords, sentences, or paragraphs that appeared meaningful and contributed to

describing the data. Relevant information was coded to reflect potential themes or sub-themes, conducted both manually and with the help of MAXQDA qualitative analysis software.

Larger themes or categories represented common patterns in the data, helping the researcher identify relationships and recurring structures. The process of grouping codes into broader categories followed these steps:

Code Identification: This step involved recognizing words, phrases, or indicators that pointed to specific themes.

Code Analysis and Review: The texts associated with each code were read carefully to identify existing patterns within the data.

Grouping Codes and Assigning Them to Themes: Codes were grouped based on semantic and conceptual similarities into broader themes.

During the grouping process, several considerations were made:

Codes assigned to a broader theme were conceptually and semantically related.

The number of codes per theme was kept reasonable and proportionate.

Themes were sufficiently developed to provide meaningful insights.

Themes aligned with the research objectives and questions.

Finally, themes and codes were analyzed meticulously to ensure accurate interpretation and to facilitate a deeper understanding of the data.

Moreover, the categorization and hierarchy used in this research for grouping themes included four main levels: sub-themes, basic themes, organizing themes, and overarching themes. The following sections provide a detailed explanation of these themes, their order, sequence, and the relationships among them.

Sub-Themes: These are subsets of basic themes. Sub-themes facilitate a deeper analysis and may include more detailed aspects of the basic themes.

Basic Themes: Positioned at the lowest level of the hierarchy, following sub-themes, basic themes are directly extracted from the data or sub-themes and contain specific details and objective information.

Organizing Themes: Located at the intermediate level of the hierarchy, organizing themes encompass several basic themes and help in identifying patterns and relationships among them. These themes support deeper data analysis and provide a structure for better understanding.

Overarching Themes: At the highest level of the hierarchy, overarching themes include multiple organizing and basic themes, contributing to a comprehensive analysis of the topic. They represent the overall findings.

In the final stage of analysis (confirmation), the process of defining, naming, and interpreting themes was carried out to enable deeper understanding and clarify the relationships among themes. Each theme was carefully defined and named to clearly represent the specific aspect of the phenomenon under study. Additionally, a concise but comprehensive description was provided for each theme to accurately explain its significance and role within the overall research framework.

Based on these steps, the results of the thematic analysis are presented in Table 3.

Table 3 – Dimensions, Components, and Indicators Related to the Validity of Designed Models in the Humanities Using a Mixed Method Approach

Dimension	Component	Indicator
External Validity	Objective	Ability of the designed model to describe the studied phenomenon
		Ability of the designed model to explain the phenomenon in real-world contexts
		Ability of the designed model to predict the phenomenon in real-world contexts
		Ability of the designed model to control the phenomenon in real-world contexts
	Research Design	Use of an appropriate research type based on classifications such as paradigm, data type, objective, etc., in model design
		Use of an appropriate sample size and sampling method to ensure representativeness of participants in model design
		Use of appropriate measurement tools and evaluation of their validity and reliability in model design
		Use of appropriate statistical methods for data analysis in model design
	Control of Confounding Variables	Ability of the designed model to control environmental conditions in terms of geographic location for generalization of findings
		Ability of the designed model to control cultural factors for generalization of findings
		Ability of the designed model to control social factors for generalization of findings
		Ability of the designed model to control political factors for generalization of findings
		Ability of the designed model to control economic factors for generalization of findings
		Ability of the designed model to control legal factors for generalization of findings
		Ability of the designed model to control religious factors for generalization of findings

		Ability of the designed model to generalize findings across different times, places, and domains
	Alignment	Results of the designed model are consistent with prior empirical studies in similar contexts
		Related research documentation clearly supports the designed model
		Results of the designed model are compatible with previous theories in similar contexts
		Theoretical foundations relevant to the topic clearly support the designed model
		Results of the designed model can be generalized to other theories in similar fields
		The proposed model has clearly defined strengths and weaknesses compared to existing models
		Results of the designed model are improved or different compared to other models
		Results of the designed model are consistent with other existing models
Internal Validity	Logical Review	The model appears logically and scientifically sound and can accurately explain relationships among its components
		Constructs are properly developed and named based on scientific principles and theoretical foundations
		Relationships among model components are clearly and logically explained and empirically evaluable
	Expert Feedback	Expert feedback clearly contributes to strengthening or revising the model
		Expert opinions assist in refining relationships among identified factors and improving accuracy
		Experts’ suggestions are consistent with their practical experience in the research field
		Expert recommendations help improve relationships among model components and enhance model validity

	Sensitivity Analysis	Small changes in model inputs do not significantly affect final results or relationships among components Critical points in the model have been identified where input changes do not have substantial effects The model effectively predicts outcomes under varying input conditions, and these predictions are stable

The table presented examines the validity of the research model from two primary dimensions: external validity and internal validity. In this context, the components associated with each dimension and the number of items corresponding to each component are specified.

1. External Validity

External validity refers to the model's ability to generalize the results to other populations and different conditions. This dimension comprises four main components, as outlined below:

Objective (4 items): This component emphasizes the clarity of the research objective and how this objective can lead to generalizable results. The presence of 4 items in this section highlights the critical importance of precisely defining the research objectives.

Research Design (4 items): This component pertains to the methods employed for data collection and analysis. An appropriately designed research methodology can enhance the external validity of the results.

Control of Confounding Variables (8 items): This component involves identifying and controlling irrelevant or extraneous factors that may influence the research outcomes. The inclusion of 8 items underscores the significance and complexity of controlling confounding variables in validating the model.

Alignment (7 items): Alignment refers to assessing the consistency between the research conditions and real-world settings. This component ensures that the research findings can be generalized across various contexts.

Internal Validity

Internal validity refers to the credibility of the results within the model and whether the findings accurately reflect reality. This dimension comprises three components, as described below:

Logical Review (3 items): This component examines the logical consistency of the assumptions and outcomes. The relatively smaller number of items in this section may indicate that logical review, as a preliminary and essential process in model validation, is generally summarized in a concise manner.

Expert Feedback (4 items): This component involves collecting opinions and evaluations from subject-matter experts. Receiving feedback from experts can help enhance the internal validity of the model.

Sensitivity Analysis (3 items): Sensitivity analysis examines how results are affected by changes in assumptions or variables. This component assists in assessing the stability of the results under various modifications.

Finally, the complete information regarding the questionnaire is presented in the table below:

Table 4 – Distribution of Questionnaire Items for Model Validation Using the Parceling Method

Questionnaire	Dimension	Component	Number of Items	Score Range	Response Scale	Item Scores
Model Validity	External Validity	Objective	4	4–20	Five-point Likert scale	5 = Very High; 4 = High; 3 = Moderate; 2 = Low; 1 = Very Low
		Research Design	4	4–20		
		Control of Confounding Variables	8	8–40		
		Adaptation	7	7–35		
	Internal Validity	Logical Review	3	3–15		
		Expert Feedback	4	4–20		
		Sensitivity Analysis	3	3–15		

To assess the validity of the designed questionnaire from the respondents' perspective, the 34-item instrument, using a five-point Likert scale (ranging from Very Low to Very High), was distributed among 357 faculty members. The results are presented below.

Initially, the factor loadings of the measurement model are shown.

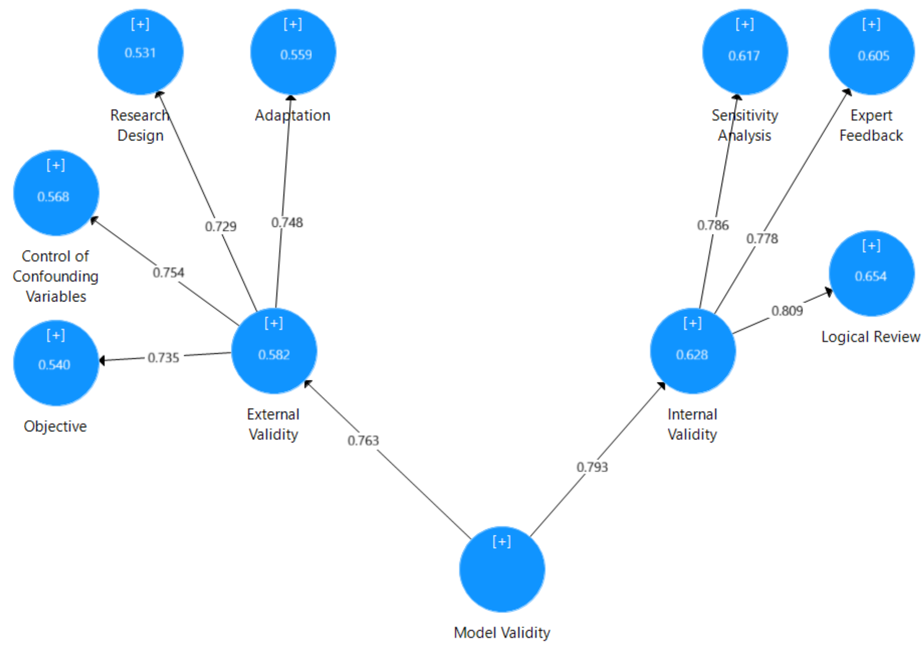

Figure 1 – Graphical Representation of Factor Loadings in the Measurement Model

As illustrated in the figure above, the factor loadings for all components and indicators exceed 0.4, indicating that the indicators adequately represent their respective components and that the components effectively reflect their corresponding dimensions.
The figure below presents the T-values for model significance.

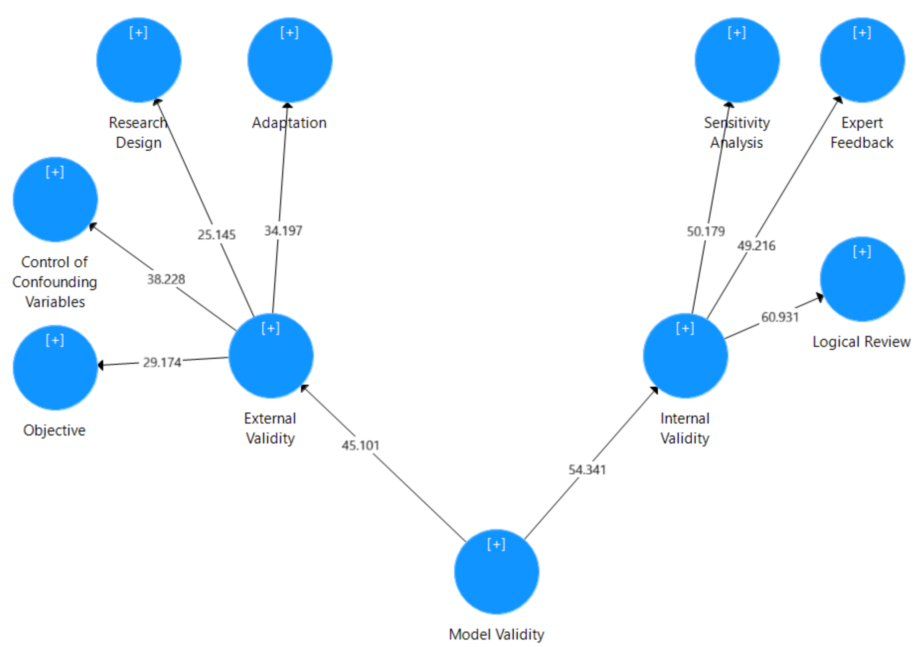

Figure 2 – Graphical Representation of Significance Coefficients in the Measurement Model

As shown in the figure above, the T-values for all indicators and components exceed 2.58, indicating that all indicators adequately represent their respective components and all components effectively reflect their corresponding dimensions with 99% confidence. Therefore, no indicators or components need to be removed from the model.

Model Fit Assessment

The purpose of evaluating the overall model fit is to determine the extent to which the entire model aligns with the empirical data. Structural Equation Modeling (SEM) combines confirmatory analysis and multivariate regression. In this approach, the overall model assessment includes two main components:

Measurement Model Assessment – examining reliability and validity of constructs.

Structural Model Assessment – evaluating path coefficients and explained variance.

R^2 (Coefficient of Determination): This index reflects the proportion of variance in endogenous (dependent) latent variables explained by independent variables. R^2 values are interpreted as follows: 0.19 = weak, 0.33 = moderate, and 0.67 = strong. In this study, R^2 values for external and internal validity were 0.582 and 0.628, respectively, indicating a strong explanatory power.

Communality Index: This measure indicates how much of the variability of the indicators (questions) is explained by their associated construct. The average communality, used to assess convergent validity, was calculated as 0.56.

Goodness-of-Fit (GoF) Index: Using the geometric mean of R^2 and the average communality for the entire model, the GoF was calculated as 0.573, indicating an acceptable fit of the model to the data.

$Gof=\sqrt{\overline{communalities} \times \overline{R^2}} = \sqrt{0.56*0.587}=0.573$

Since the calculated Goodness-of-Fit (GoF) for the primary research model exceeds 0.36, this indicates a satisfactory fit of the model.

Predictive Relevance (Q^2): This index assesses the model's predictive power for the dependent variables. The interpretation criteria for Q^2 are 0.02 = small, 0.15 = medium, and 0.35 = large predictive power. In this study, Q^2 was 0.223, indicating an acceptable level of predictive capability for the model's variables.

Normed Fit Index (NFI): The NFI, a non-normed fit index, evaluates how well the proposed model improves fit relative to a null model, with values greater than 0.9 considered acceptable. For this model, NFI = 0.966, demonstrating excellent model fit.

Based on these findings, it can be concluded that the tested model exhibits a good fit for the examined sample. Additionally, since all factor loadings of the observed variables exceed 0.4 and their t-values are greater than 1.96, the constructs demonstrate satisfactory validity.

Discussion and conclusion

The findings of the study indicate that the designed 34-item questionnaire serves as an appropriate instrument for assessing the validity of models developed in the humanities from the perspective of experts. This questionnaire comprehensively evaluates the external validity of models through four components: Goal (4 items), Research Design (4 items), Control of

Confounding Variables (8 items), and Alignment (8 items). Additionally, the internal validity of the models is assessed through three components: Logical Review (3 items), Expert Feedback (4 items), and Sensitivity Analysis (3 items).

The inclusion of these components ensures coverage of all the main dimensions related to model validation, providing researchers with the ability to conduct precise and scientifically robust evaluations of the models.

The design and validation of models in the humanities have always been challenging due to the complex and multifaceted nature of these disciplines. A review of theoretical foundations and existing literature in reputable databases such as Scopus and PubMed indicates that external and internal validity are among the primary criteria used by researchers to evaluate models across various studies.

For instance, William (2024) demonstrated that controlling confounding variables and the precise design of research methods are key factors in enhancing the reliability of scientific models. These findings align closely with the external validity components of the questionnaire developed in the present study, as these components ensure that the models are capable of producing valid results not only theoretically but also practically.

Furthermore, Lim (2024) highlighted the importance of logical review and expert feedback in model validation. As elements of internal validity, these aspects play a crucial role in assessing the quality of model design. Logical review, by focusing on a detailed analysis of the model components, can prevent errors in the early stages of design, while expert feedback enriches the models and enhances their credibility.

In international studies, Asher (2024) emphasized the importance of sensitivity analysis in examining model stability. Sensitivity analysis allows researchers to assess the impact of minor changes in model variables on overall outcomes. This component, incorporated into the questionnaire developed in the present study, ensures that the models remain applicable and valid under various conditions.

Finally, a review of the research background indicates that the integration of external and internal validity components into a comprehensive instrument constitutes an effective step toward enhancing the quality of models developed in the humanities. Similar studies, such as the research conducted by Pidrahita et al. (2025), have also emphasized the necessity of comprehensive tools for model validation, which confirms the approach adopted in the present study.

These findings underscore the importance of a systematic approach to validating models in the humanities. The 34-item questionnaire developed in this study, by covering key components of both external and internal validity, allows for the precise identification of the strengths and weaknesses of models, enabling researchers to design models with high applicability and generalizability. Given the inherently complex nature of the humanities, which requires multidimensional evaluation, this questionnaire can serve as a standardized tool for assessing model quality in this domain.

Moreover, by employing a mixed-method approach, the questionnaire facilitates a more comprehensive analysis. The combination of quantitative and qualitative components in the instrument enables researchers to evaluate models from multiple perspectives and utilize the results to improve the design of future models. This tool not only enhances the scientific validity of models but also contributes to the advancement of methodology in the humanities.

Ultimately, the findings of this study are consistent with the results of Karnia (2024), Lim (2024), Bernhard-Harrer et al. (2025), and Vahdani et al. (1404). All of these studies emphasize the importance of model validation using components such as control of confounding variables, logical review, expert feedback, and sensitivity analysis. The alignment of these components with the questionnaire developed in the present research demonstrates the validity and comprehensiveness of the proposed tool and confirms its applicability for the scientific evaluation of models developed in the humanities.

Based on the findings of the study, the following practical recommendations are proposed. These suggestions are directly derived from the research results and address the challenges and needs of key stakeholders. They are concrete, actionable, and grounded in scientific and empirical evidence:

1. Designing and Implementing Training Programs for Humanities Researchers

Description of the Recommendation: Organize specialized workshops for humanities researchers on the use of the 34-item questionnaire developed in this study. These workshops should provide practical training on how to design valid models and assess their validity effectively.

Implementation Steps:

• Scheduling: Plan and conduct three-month-long workshops throughout the academic year.

• Required Resources: Prepare educational materials, including a comprehensive guide on using the questionnaire and practical examples.

• Responsibilities: University faculty members to serve as workshop instructors; research groups within universities to organize logistics and coordination.

Evaluation and Monitoring: Collect feedback forms from participants to assess the quality of the workshops and the practical applicability of the training.

Positive Impacts: This recommendation can enhance the skills of researchers in designing valid models and improve the overall quality of research in the humanities.

2. Development of Reliable Databases for Storing and Sharing Designed Models

Description of the Recommendation: Establish a national database for storing designed models in the humanities that have been validated using the 34-item questionnaire. This database can serve as a reference for researchers and policymakers.

Implementation Steps:

• Scheduling: Design and launch the database within six months.

• Required Resources: Allocate a budget for IT infrastructure development and hire IT specialists.

• Responsibilities: Ministry of Science, Research, and Technology to provide funding; universities to contribute validated models.
Evaluation and Monitoring: Annually assess the level of researchers' utilization of the database and the quality of stored models.
Positive Impacts: This database can facilitate knowledge sharing, enhance collaboration among humanities researchers, and provide stakeholders with access to validated models.

3. Establishing a Framework for Interdisciplinary Collaboration in Model Design and Validation
Description of the Recommendation: Form interdisciplinary research groups to design and evaluate humanities models using the validated questionnaire. These groups can include experts from humanities, statistics, social sciences, and data science.
Implementation Steps:
• Scheduling: Form research groups within universities over three months and implement collaborative projects over the course of one year.
• Required Resources: Allocate funding for interdisciplinary projects and facilitate communication among research groups.
• Responsibilities: Universities and research centers to establish the groups; Ministry of Science to provide financial support.
Evaluation and Monitoring: Assess the quality of completed projects and the level of interdisciplinary interaction.
Positive Impacts: This recommendation can enhance interdisciplinary knowledge, promote the development of more comprehensive and applicable models, and strengthen collaborative research in the humanities.
The present study, despite its significant scientific contributions, has certain limitations that may affect the validity, generalizability, and applicability of the findings. These limitations include:
1. Limitation in the Statistical Population: The quantitative phase of the study included only faculty members and professors from humanities disciplines at universities in Tehran. This geographical limitation may affect the generalizability of the results to other cities or academic communities outside of Iran.
2. Impact of Sampling Methods: The use of purposive non-random sampling in the qualitative phase and multi-stage cluster sampling in the quantitative phase may result in findings that do not fully represent the broader statistical population.
To address these limitations and enhance the quality of future research, the following recommendations are proposed:
1. Expanding the Statistical Population: Future researchers can extend the study population to include faculty members and researchers in the humanities from other cities in Iran or even from other countries. This expansion would increase the generalizability of the findings and enable comparisons across different cultures and academic communities.

2. Utilizing Random Sampling Methods: To enhance the validity of results, it is recommended to employ random sampling methods in both the qualitative and quantitative phases. This approach can provide a more accurate representation of the overall population and improve the statistical robustness of the findings.
3. Establishing Interdisciplinary Studies: It is suggested that future researchers collaborate with experts from related fields such as statistics, data science, and others to design more comprehensive validation models. Such interdisciplinary collaborations can lead to the development of validation tools with broader applicability and greater methodological rigor.

References

Ansarian Tehrani, A. (2025). Providing a framework for designing interactive MOOCs in higher education in Iran with an emphasis on humanities disciplines (Faculty of Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran).

Bandani Pour, M. (2024). Developing a financial fraud model with emphasis on political factors and the diamond and fraud triangle components (Doctoral dissertation, Faculty of Humanities, Birjand Branch, Islamic Azad University, Birjand, Iran).

Parafkandeh Haghighi, M. (2025). Developing a model for enhancing educational justice for teachers in the counties of Tehran province (Faculty of Management, Tehran West Branch, Islamic Azad University, Tehran, Iran).

Hossein Pour, S. (2024). Designing a model for evaluating the implementation of transparency policies in Iranian public organizations: A case study of the Ministry of Industry, Mine, and Trade (Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran).

Hossein Zadeh, M. (2024). Developing a model for identifying and enrolling out-of-school children (primary level) with a human development approach in Iran (Faculty of Psychology and Educational Sciences, Central Tehran Branch, Islamic Azad University, Tehran, Iran).

Darabian, H. (2025). Developing a model to enhance competencies required for the future labor market: A new mission of Islamic Azad University (Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran).

Shadalouei, M. (2025). Developing a model of factors affecting social demand for higher education in Iran (Faculty of Educational Sciences, South Tehran Branch, Islamic Azad University, Tehran, Iran).

Ghasemian, H. (2025). Developing a model for managerial sense-making in schools to implement the Fundamental Transformation Document of Education (Faculty of Management, Rudehen Branch, Islamic Azad University, Rudehen, Iran).

Asher, A. D. (2024). Approaches for Improving Validity in Quantitative Research Articles. portal: Libraries and the Academy, 24(2), 209-215.

Bernhard-Harrer, J., Ashour, R., Eberl, J. M., Tolochko, P., & Boomgaarden, H. (2025). Beyond standardization: a comprehensive review of topic modeling validation methods for computational social science research. Political Science Research and Methods, 1-19.

Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. Psychological Inquiry, 31(4), 336-344.

Karnia, R. (2024). Importance of reliability and validity in research. Psychology and Behavioral Sciences, 13(6), 137-141.

Kulesa, J., Induru, S., Hubbard, E., & Bhansali, P. (2024). The Conceptual Framework: A Practical Guide. Hospital pediatrics, 14(11), e503-e508.

Lim, W. M. (2024). A typology of validity: content, face, convergent, discriminant, nomological and predictive validity. Journal of Trade Science, 12(3), 155-179.

Lorenzini, E., Osorio-Galeano, S. P., Schmidt, C. R., & Cañon-Montañez, W. (2024). Practical guide to achieve rigor and data integration in mixed methods research. Investigación y Educación en Enfermería, 42(3).

Piedrahita, A. R., Ardila, L. M. C., Valencia, J. A. P., Dores, Á. J. L. D., & Canencio, J. G. P. (2025). Modeling and Simulation with System Dynamics: A Literature Review on Validation Aspects. Journal of Hunan University Natural Sciences, 52(1).

Schoonenboom, J. (2024). Design patterns in mixed methods research: Embedding mixed methods matched comparisons sampling in previous design decisions. Journal of Mixed Methods Research, 18(3), 383-392.

Varpio, L., Paradis, E., Uijtdehaage, S., & Young, M. (2020). The distinctions between theory, theoretical framework, and conceptual framework. Academic medicine, 95(7), 989-994.

William, F. K. A. (2024). Mastering validity and reliability in academic research: Meaning and significance. International Journal of Research Publications, 144(1), 287-292.