Scilight

Journal of Modern Applied Statistical Methods
https://jmasm.com

Article

Rank-Based Item-Level Transformation for Harmonizing Multi-Scale Likert Data: A Comparative Evaluation

Carl Beitsayadeh [1,2] and Pamayla E. Darbyshire [2,*]

[1] College of General Studies, University of Phoenix, Phoenix, AZ 85040-0723, USA
[2] College of Doctoral Studies, University of Phoenix, Phoenix, AZ 85040-0723, USA
* Correspondence: mashg@email.phoenix.edu

How To Cite: Beitsayadeh, C.; Darbyshire, P.E. Rank-Based Item-Level Transformation for Harmonizing Multi-Scale Likert Data: A Comparative Evaluation. *Journal of Modern Applied Statistical Methods* **2026**, *25*(2), 8. https://doi.org/10.53941/jmasm.2026.100008

Received: 26 February 2026
Revised: 7 April 2026
Accepted: 8 June 2026
Published: 2 July 2026

Abstract: Integrating Likert-type data from multiple instruments with heterogeneous response scales presents a persistent methodological challenge in applied multivariable modeling contexts involving ordinal data. This quantitative secondary analysis evaluates a rank-based transformation procedure, the Rank-Based Harmonization Framework (RBHF), alongside five benchmark transformations applied at the item level prior to aggregation. Methods were compared using descriptive distributional statistics, correlational analysis, multivariable linear regression, and residual diagnostics. Results indicated measurable differences across transformation strategies. Relative to benchmark methods, RBHF produced construct scores exhibiting closer alignment with normality indicators, smaller deviations from baseline correlation structures, comparatively larger adjusted R^2 values, and comparatively stable residual variance patterns within the analytic conditions examined. These findings reflect observable variation in how transformation procedures influence distributional properties, relational fidelity, and regression diagnostics when harmonizing multi-scale ordinal data. The RBHF procedure implements a minimum-rank inverse normal transformation at the item level and may be useful in applied research contexts where heterogeneous Likert-type instruments must be harmonized for joint analysis within parametric modeling frameworks.

Keywords: data harmonization; Likert-scale transformation; multivariable regression; ordinal data analysis; rank-based inverse normal transformation; survey methodology

1. Introduction

In applied research settings across the social and behavioral sciences, Likert-type instruments remain widely used tools for capturing perceptions, attitudes, and behaviors related to learning, motivation, and engagement. These instruments consist of multiple ordinal-response items, typically ranging from three to seven response categories, quantifying participants' experiences in structured formats [1–4]. While such tools offer diagnostic value and practical measurement advantages, combining data from multiple instruments with heterogeneous and multi-scale measurement structures introduces persistent methodological challenges.

A common practice in multivariable educational analyses involves integrating survey data from two or more Likert-type instruments, each with its own scale length and response characteristics. Although each instrument may be individually validated, discrepancies in their scale formats can introduce statistical inconsistencies when variables are modeled jointly. Differences in range, midpoint semantics, and item anchoring can bias correlation matrices, distort regression coefficients, and undermine the assumptions required for parametric analysis, most notably normality, linearity, and homoscedasticity [5–8].

These problems are particularly pronounced in applied multivariable analyses where constructs are measured using heterogeneous ordinal instruments. For example, a 5-point Likert scale designed to assess behavioral engagement may be analyzed alongside a 4-point instrument developed to measure affective disengagement. Without appropriate harmonization techniques, this structural heterogeneity limits valid cross-instrument comparison and inflates error variance in multivariable modeling frameworks [9–11].

To address this persistent issue, this methodological analysis introduces and evaluates the Rank-Based Harmonization Framework (RBHF), a novel item-level transformation method designed to normalize ordinal data while preserving its rank-order structure and improving compatibility with parametric techniques. Unlike benchmark transformation methods that apply average-ranking or distribution-driven adjustments, RBHF employs a minimum-rank inverse normal transformation at the item level prior to aggregation, ensuring that strict ordinal relationships are maintained even in the presence of tied values. This approach differs from traditional methods by prioritizing rank preservation first, then enhancing statistical conformity through transformation [12–15]. For fair methodological comparison, all benchmark methods evaluated (z-score standardization, inverse normal transformation (INT), Rankit, quantile normalization, and raw linear rescaling) were also applied at the item level prior to aggregation.

A distinctive feature of the present analysis is its inclusion of an instrument with a nonstandard 4-point Likert scale, originally designed for general-purpose use. The instrument set included a range of Likert-type formats, including general-purpose tools with nonstandard scale structures and polarity patterns, further underscoring the need for robust harmonization strategies [16–18]. RBHF's ability to accommodate these diverse formats illustrates its potential applicability across varying scale structures. By operating at the item level and preserving rank-order structure, RBHF is well positioned to address the analytical demands of multivariable modeling involving heterogeneous ordinal instruments.

The purpose of this methodological analysis is to evaluate and compare the performance of six item-level transformation techniques applied to ordinal data from multiple Likert-type instruments, with a focus on statistical assumptions, relational fidelity, and predictive modeling. This comparison addresses a persistent methodological challenge in educational and psychological research [19–21].

The empirical dataset used for this comparison was drawn from a previously published survey study of online higher education students and was repurposed here as an applied methodological test case.

Transforming ordinal data to better approximate normality is often necessary because many commonly used parametric statistical techniques, including multivariable linear regression, rely on assumptions of normality, linearity, and homoscedasticity for valid inference. Failure to satisfy these assumptions may lead to biased parameter estimates, reduced statistical power, and compromised interpretability [22,23].

The findings are intended to contribute to methodological decision-making in analyses involving multi-scale ordinal data within multivariable parametric modeling contexts. In addition to evaluating each criterion independently, a composite performance index was constructed post hoc to summarize comparative performance across distributional, correlational, predictive, and residual domains. The composite index was constructed as a descriptive aggregation tool to summarize relative performance patterns rather than to infer population-level superiority [24–26].

2. Literature Review

2.1. The Nature of Likert Scales in Applied Research

Likert-type items remain foundational in educational and psychological research due to their structured simplicity, cost-effectiveness, and diagnostic utility in measuring latent variables such as attitudes, motivation, and engagement. Originally introduced [1], these ordinal-scale items allow respondents to express varying degrees of agreement or frequency across ordered response categories. Although inherently ordinal, these data are often analyzed using parametric statistical techniques under the assumption of underlying continuity [4,7,27].

In educational survey research, multiple instruments with differing Likert formats, ranging from 3-point to 7-point scales, are frequently employed within a single analysis to assess conceptually distinct variables. This structural heterogeneity introduces methodological challenges, particularly when attempting to compare or aggregate data across instruments with divergent scale structures [28,29].

Recent psychometric literature highlights ongoing concerns about treating Likert-type responses as interval-level data without appropriate transformation. Robitzsch (2020) [27] argues that while ordinal data can often be treated as continuous for analytical purposes, this practice must be accompanied by methodological safeguards to ensure validity. Similarly, Kusmaryono et al. (2022) [18] emphasize the importance of scale reliability and caution against simplistic interpretations that overlook structural differences among instruments. Without appropriate safeguards, treating ordinal data as continuous can result in biased regression estimates, invalid significance tests, and misleading substantive interpretations.

A central issue is scale non-equivalence: for example, a mid-scale rating of "3" on a 4-point scale does not map linearly or semantically to a "3" on a 7-point scale. Without appropriate transformation techniques, such inconsistencies can introduce statistical bias, distort correlational and regression estimates, and undermine the validity of model-based inferences [28,29].

2.2. Challenges in Cross-Instrument Comparison

The use of multi-instrument survey designs, while conceptually robust, introduces significant analytical complexity when the instruments employ different Likert-type scale formats. Discrepancies in the number of response categories, item polarity, midpoint semantics, and anchoring definitions can compromise statistical integrity and interpretive clarity. As Carifio and Perla (2008) [9] argue, and as later reaffirmed by Kankaraš et al. (2018) [30], such inconsistencies violate assumptions of interval-level equivalence, potentially inflating measurement error or obscuring true relationships among variables. More broadly, the misuse of ordinal Likert responses in statistical analyses has also been highlighted as a persistent methodological concern, particularly when measurement assumptions are overlooked [31].

Even when ordinal data are treated as continuous, a practice defended by Norman (2010) [7] for its robustness in simple analyses, applied researchers must remain cautious of challenges such as skewed distributions, ceiling and floor effects, and heteroscedastic residuals, which become especially pronounced in multivariable models involving multiple ordinal variables measured on differing Likert-type scales. Rhemtulla et al. (2012) [5] demonstrate that while continuous estimation methods (e.g., normal theory maximum likelihood) can perform adequately when ordinal scales have five or more categories and category thresholds are symmetric, violations of key assumptions like linearity, normality, and homoscedasticity can still occur. These issues are particularly problematic in multivariable models where multiple variables measured on multi-scale ordinal instruments are analyzed simultaneously. Violations of statistical assumptions not only affect model fit but also weaken the reliability and validity of inferences drawn from parameter estimates and hypothesis tests.

Accordingly, robust transformation procedures are essential to ensure comparability, preserve theoretical constructs, and enable accurate statistical modeling across instruments with heterogeneous scale characteristics.

2.3. Prior Transformation and Harmonization Approaches

A variety of data transformation techniques have been applied to prepare ordinal Likert-type responses for use in parametric models. The most common method is linear rescaling, which maps responses from different scales onto a shared range, typically 0 to 100. This approach preserves ordinal spacing and enables visual comparability but does not address issues of non-normality or unequal item spacing inherent in ordinal data [4].

Recent methodological work has further emphasized the importance of respecting ordinal response structure in analytical decision-making, particularly for polytomous data where item-level properties influence downstream statistical behavior [20,21].

Blanca et al. (2017) [8] demonstrate that real psychological data, including those from Likert-type instruments, frequently violate the assumption of normality, which underlies the use of z-score standardization and other parametric techniques. Z-score standardization is a widely used technique that normalizes the mean and standard deviation of variables, enabling direct comparisons across instruments. However, this method relies on the assumption of interval-level measurement and normally distributed data, conditions that are often violated in Likert-type scales. Jamieson (2004) [2] explicitly critiques the misuse of parametric statistics (e.g., means, standard deviations, and ANOVA) for ordinal Likert data, arguing that such practices are illegitimate because the intervals between scale points cannot be presumed equal. She further highlights that Likert-derived data frequently exhibit skewed or polarized distributions, undermining the normality assumption required for parametric techniques. While z-scores are not mentioned directly in her article, her broader warnings about violating measurement assumptions and misapplying parametric methods to ordinal data implicitly challenge the validity of z-score standardization for Likert scales.

Rank-based transformations, such as the inverse normal transformation (INT) and Rankit [12,13], offer a nonparametric alternative by mapping rank-ordered responses to z-scores based on cumulative probability. These methods are more suitable for ordinal or non-normal data, as they reduce skew and improve conformity to statistical assumptions [32,33]. However, they are often applied only after aggregation at the variable level, which discards item-level variance and weakens interpretability.

Quantile normalization, more common in genomic and machine learning contexts, forces distributions to match a common reference. While effective in aligning variable distributions, this method can distort rank order and obscure theoretical meaning when applied to social science data [34].

Existing transformation approaches offer partial solutions but tend to prioritize statistical conformity over conceptual fidelity. Most existing approaches do not simultaneously preserve item-level ordinal structure, retain inter-individual variance, and satisfy the assumptions required for valid parametric modeling. These methodological trade-offs result in gaps that can compromise both interpretability and statistical validity in multivariable analyses, gaps that the Rank-Based Harmonization Framework is explicitly designed to address.

2.4. Theoretical Rationale for Harmonization

Efforts to harmonize ordinal data from multiple Likert-type instruments are grounded in the theoretical imperative to ensure interpretive coherence across diverse measurement tools. When instruments are used to assess related but distinct variables, each with its own scale structure, theoretical constructs can become obscured if scale-induced artifacts distort their statistical relationships. Harmonization techniques are therefore necessary not only to facilitate model compatibility but also to preserve the integrity of conceptual interpretations across variables [20,21].

From a measurement theory perspective, harmonization involves reconciling the ordinal nature of Likert-type data with the linear assumptions of common statistical models. Traditional scale theory treats ordinal responses as reflecting latent continuous traits, but without proper transformation, this approximation risks violating core assumptions such as linearity, normality, and homoscedasticity [5]. In multivariable models where variables measured on different scales are examined concurrently, these violations can undermine parameter accuracy, inflate error variance, and reduce the overall validity of statistical inferences.

Beyond statistical assumptions, harmonization also serves an important theoretical role by enabling comparability across instruments that differ not only in format but also in response semantics. For example, scales anchored by “Never” to “Always” may imply different interpretive intensities than those using “Strongly Disagree” to “Strongly Agree,” even if they contain the same number of points. Differences in response scale formats can influence response styles and interpretation, potentially affecting the comparability of measurement outcomes across instruments [29].

Accordingly, this methodological analysis treats harmonization not merely as a preprocessing step, but as a theoretically necessary intervention that enables robust and meaningful analysis of ordinal data drawn from multiple heterogeneous instruments.

2.5. The Need for a New Framework

This methodological analysis contributes to ongoing efforts in educational and social sciences research to harmonize ordinal data across structurally divergent instruments. Unlike prior approaches that often prioritize statistical normalization at the expense of conceptual coherence, the method introduced and evaluated here, the Rank-Based Harmonization Framework (RBHF), balances both objectives by applying a minimum-rank inverse normal transformation at the item level prior to aggregation. This approach builds upon established rank-based transformation methods [12–15,32] while extending their application to item-level harmonization across heterogeneous Likert-type instruments. By applying transformation prior to aggregation, the framework retains ordinal structure while simultaneously supporting parametric modeling assumptions.

What distinguishes the Rank-Based Harmonization Framework is its ability to accommodate diverse response formats, including general-purpose instruments with nonstandard scaling, without introducing interpretive distortion. By operating at the item level and relying on a minimum-rank transformation scheme, the framework preserves rank-order fidelity and enhances variable comparability across instruments, even when polarity and scale semantics differ.

In addition to evaluating the statistical behavior of six transformation methods across four independent criteria, distributional shape, relational fidelity, predictive strength, and residual behavior, this analysis also introduces a composite performance index. Composite evaluation approaches are increasingly used in methodological comparisons to summarize performance across multiple statistical criteria [24–26]. This index provides a unified metric of overall methodological utility, enabling researchers to select transformation approaches that are not only statistically robust but also theoretically aligned with the interpretive demands of multivariable models. Taken together, these contributions support improved practices in harmonizing multi-scale ordinal data and offer an adaptable, scalable framework for future applications involving cross-instrument survey designs.

3. Research Purpose and Questions

3.1. Purpose of the Methodological Analysis

The primary purpose of this methodological analysis was to evaluate and compare the effectiveness of six data transformation techniques in harmonizing Likert-type responses collected through multiple instruments employing heterogeneous and multi-scale ordinal scale structures. In applied survey research, instruments frequently adopt differing Likert formats, such as 4-point, 5-point, and 7-point response options, complicating valid cross-variable comparison and multivariable statistical modeling.

This analysis introduced and assessed a newly proposed transformation method, the Rank-Based Harmonization Framework, alongside five established benchmark methods: raw linear rescaling (0–100), z-score standardization, inverse normal transformation (INT), Rankit transformation, and quantile normalization. All six methods were applied at the item level prior to aggregation.

The overarching objective was to examine how the evaluated methods support parametric assumptions of normality and homoscedasticity (e.g., skewness and kurtosis), preserve rank-order relationships, preserve correlational relationships among constructs, and influence regression model performance (e.g., R^2 values, standardized β coefficients, and residual behavior). The findings are intended to inform the selection of transformation techniques for researchers analyzing heterogeneous ordinal survey data in multivariable modeling contexts where scale non-equivalence presents methodological challenges.

3.2. Research Questions and Hypotheses

To guide the methodological comparison, four research questions were formulated addressing key statistical properties relevant to harmonizing heterogeneous Likert-type data in multivariable analyses.

Research Question 1 (RQ1). *How do the evaluated transformation methods differ in their ability to improve distributional properties of Likert-type construct scores derived from heterogeneous ordinal instruments*?

Null Hypothesis ($\mathbf{H_{01}}$). *The evaluated transformation methods do not differ in their effects on distributional properties of the resulting construct scores, as assessed by indicators such as skewness and kurtosis*.

Alternative Hypothesis ($\mathbf{H_{A1}}$). *The evaluated transformation methods differ in their effects on distributional properties of the resulting construct scores, as assessed by indicators such as skewness and kurtosis*.

Research Question 2 (RQ2). *To what extent do the transformation methods preserve the correlational relationships among the constructs*?

Null Hypothesis ($\mathbf{H_{02}}$). *The evaluated transformation methods do not differ in the degree to which they preserve the original correlational structure among constructs, as assessed by deviations from baseline correlation coefficients*.

Alternative Hypothesis ($\mathbf{H_{A2}}$). *The evaluated transformation methods differ in the degree to which they preserve the original correlational structure among constructs, as assessed by deviations from baseline correlation coefficients*.

Research Question 3 (RQ3). *How do regression models derived from each transformation method differ in terms of model performance and diagnostic behavior*?

Null Hypothesis ($\mathbf{H_{03}}$). *The evaluated transformation methods do not differ in regression model performance, including explained variance (R^2), standardized regression coefficients, and residual diagnostics*.

Alternative Hypothesis ($\mathbf{H_{A3}}$). *The evaluated transformation methods differ in regression model performance, including explained variance (R^2), standardized regression coefficients, and residual diagnostics*.

Research Question 4 (RQ4). *Which transformation method demonstrates the most balanced overall performance across distributional properties, correlational fidelity, regression model performance, and residual diagnostics*?

Null Hypothesis ($\mathbf{H_{04}}$). *No transformation method demonstrates a meaningfully different overall performance profile across the evaluated analytical domains*.

Alternative Hypothesis ($\mathbf{H}_{\mathbf{A4}}$). *At least one transformation method demonstrates a meaningfully different overall performance profile across the evaluated analytical domains.*

$$nr_i^{transformed} r_i^{raw}$$

4. Methodology

4.1. Research Design

This methodological analysis employed a quantitative, comparative, secondary data analysis design to evaluate the performance of six transformation techniques applied to Likert-type survey responses. The primary objective was to comparatively evaluate how the transformation methods harmonize ordinal data while supporting the statistical assumptions required for valid multivariable modeling.

Data Source

The dataset used in this methodological analysis was derived from a previously published cross-sectional survey study examining factors influencing student academic outcomes in online higher education [35]. The original study investigated relationships among self-directed learning, collaborative learning, isolation, academic achievement, and academic resilience using validated Likert-type instruments administered to undergraduate students enrolled in online courses. The dataset contained responses from 561 participants, with a final cleaned analytic sample of 268 complete cases after data screening procedures described in the original study. The survey instruments employed heterogeneous Likert-type scales including 4-point, 5-point, and 7-point response formats. These heterogeneous ordinal structures provided an appropriate empirical context for evaluating data harmonization procedures designed for multivariable statistical modeling. In the present manuscript, the dataset is used exclusively as an applied methodological test case to examine the comparative behavior of six transformation methods applied to heterogeneous ordinal data. The objective of this analysis is therefore methodological rather than substantive, focusing on how alternative transformation procedures influence distributional properties, correlational structure, regression performance, and residual diagnostics when harmonizing multi-scale Likert data.

The demographic characteristics of the analyzed sample are described in the source study. Briefly, the sample (n = 268) consisted predominantly of female students (88.06%), with most participants between 25 and 54 years of age. Participants represented multiple academic colleges within the university, with the largest representation from Social and Behavioral Sciences (42.91%), followed by Health Professions (24.26%) and Education (20.15%). A majority of respondents identified as first-generation college students (63.43%), and most reported full-time employment while enrolled in their academic programs (58.96%). These characteristics reflect the typical demographic profile of adult learners enrolled in online undergraduate programs.

The dataset was selected for the present methodological analysis because it contains multiple constructs measured using heterogeneous Likert-type scales, providing an appropriate empirical context for evaluating transformation procedures designed to harmonize ordinal survey data in multivariable statistical modeling [35].

The survey instruments measured five constructs commonly examined in research on online learning environments: self-directed learning (SDL), collaborative learning (COL), isolation (ISO), academic achievement (AA), and academic resilience (AR). These constructs were assessed using validated Likert-type instruments employing heterogeneous response scales (4-point, 5-point, and 7-point formats). Because these instruments differ in scale structure and response range, the dataset provides an appropriate empirical context for evaluating transformation procedures designed to harmonize ordinal measurements prior to multivariable statistical analysis.

Table 1 summarizes the transformation methods evaluated in the present methodological analysis. Each transformation was applied at the item level prior to construct aggregation, allowing direct comparison of how alternative procedures influence distributional properties, correlational relationships, regression performance, and residual diagnostics when heterogeneous Likert-type data are used in multivariable statistical models.

Table 1. Overview of transformation methods evaluated.

Method	Core Procedure	Statistical Assumptions Addressed	Applied at
Raw (Rescaled to 0–100)	Linear rescaling of ordinal scores	None; preserves ordinal nature	Item level
Z-Score Standardization	Standardization to mean = 0, SD = 1	Homoscedasticity, comparability	Item level
INT (Inverse Normal Transformation)	Van der Waerden (1952) [12] method via rank probabilities	Normality, parametric test assumptions	Item level
Rankit	Rank-based z-score using Blom's correction	Variance stability, approximate normality	Item level
Quantile Normalization	Distribution alignment using rank order	Cross-variable comparability	Item level
RBHF	Minimum-rank inverse normal transformation (4-stage)	Normality, homoscedasticity, rank preservation	Item level

Note. All transformation methods were applied at the item level following initial rescaling of ordinal responses to a 0–100 metric. One instrument assessing isolation employed a 4-point ordinal response scale requiring a validated rescaling procedure provided by the instrument developers. This procedure proportionally mapped response categories to a 0–100 scale using a bounded linear transformation. These rescaled values were used as inputs for all subsequent transformation methods.

4.2. Instruments and Scaling Characteristics

Five previously validated Likert-type instruments were used to measure the constructs included in the present methodological analysis. These instruments were selected from the source study because they represent conceptually distinct domains relevant to online learning environments and employ heterogeneous ordinal response formats. Specifically, three instruments measured predictor constructs associated with learning behavior and student experience, namely self-directed learning, collaborative learning, and isolation, whereas two instruments measured academic outcome constructs, namely academic achievement and academic resilience. The instruments differed in scale length, response semantics, and scoring structure, creating the heterogeneous measurement conditions that motivated the harmonization procedures evaluated in this analysis.

The constructs were measured using the Self-Directed Learning Scale (SDLS), the Online Collaborative Learning Scale (OCLS), the UCLA Loneliness Scale 8 item version (ULS-8), the Subjective Academic Achievement Scale (SAAS), and the Academic Resilience Scale (ARS-30), as described in the source study [35]. These instruments are administered using Likert-type response formats with varying numbers of response categories, including 4-point, 5-point, and 7-point scales. Such variation in ordinal scale structure is common in applied survey research but introduces challenges when variables measured on different scales are analyzed jointly in multivariable statistical models.

To enable cross-instrument comparability, all item responses were transformed to a common 0–100 metric prior to the application of the transformation methods evaluated in this study. For four of the instruments (SDLS, OCLS, SAAS, and ARS-30), responses were rescaled using a standard linear min–max transformation based on the original Likert scale range (e.g., 1–5 or 1–7). This normalization preserved the ordinal spacing of response categories while aligning the instruments on a common numerical scale.

The exception was the Isolation scale (ULS-8), a 4-point instrument whose validated scoring procedure was developed by Hays and DiMatteo (1987) [36]. In accordance with the validated scoring procedure recommended by the instrument developers, the average raw score was transformed to a 0–100 scale using the bounded transformation:

$$(\text{raw averge score} - 1) * \left(\frac{100}{3}\right) \quad (1)$$

This transformation preserves the intended gradation between response categories (e.g., "Never" to "Often") while aligning the scale with the common 0–100 metric used for the other instruments. Applying the validated transformation ensured that the semantic interpretation of the isolation measure was retained while enabling harmonization with the other multi-scale ordinal instruments included in the dataset.

Table 2 summarizes the original scale properties and preprocessing procedures applied to each instrument prior to the implementation of the transformation methods.

Table 2. Instrument properties and preprocessing summary.

Construct/Instrument Name	Original Likert Scale	Transformation Preprocessing
Self-Directed Learning/SDLS	1–5	Rescaled to 0–100 using linear transformation.
Collaborative Learning/OCLS	1–7	Rescaled to 0–100 using linear transformation.
Isolation/ULS-8	1–4	Rescaled to 0–100 using the validated bounded transformation provided by the instrument developers.
Academic Achievement/SAAS	1–5	Rescaled to 0–100 using linear transformation.
Academic Resilience/ARS-30	1–5	Rescaled to 0–100 using linear transformation.

Note: All five instruments were administered to the full sample. The Isolation scale (ULS-8) was preprocessed using the validated bounded transformation recommended by the instrument developers prior to the application of the harmonization methods. The analytical roles of the constructs as predictor or outcome variables are described in Section 5.3 (Regression Modeling).

4.3. Data Transformation Procedures

Prior to comparative transformation, each item response from the five instruments was converted to a 0–100 scale. This preprocessing ensured that all transformation methods operated on a consistent metric while retaining ordinal integrity across the multi-scale ordinal instruments.

For four instruments, item responses were linearly rescaled using min–max normalization based on each instrument's original Likert scale range (e.g., 1–5, 1–7). The Isolation scale (ULS-8) used a nonstandard 4-point response format and was therefore rescaled using the validated bounded transformation provided by the instrument developers.

Following rescaling, all six transformation methods were applied at the item level prior to aggregation. For each respondent, composite variable-level scores were calculated by averaging the transformed item-level values within each instrument.

4.3.1. Raw (Rescaled 0–100)

For the raw condition, each item was rescaled to a 0–100 scale, and composite variable-level scores were calculated by averaging the rescaled item values within each instrument. This method preserves ordinal interpretation but does not address issues such as skewness or tied ranks.

Each item was rescaled to a 0–100 scale using:

$$\text{Item}_{(0-100)} = 100 \times \left(\frac{\text{Item Score} - \text{Minimum Value}}{\text{Scale Range}}\right) \tag{2}$$

Example for the Self-Directed Learning scale (1–5 scale):

$$= 100 \times \left(\frac{\text{Item Score} - 1}{4}\right) \tag{3}$$

For the Isolation scale (ULS-8; 4-point scale), items were rescaled using the validated formula from Hays & DiMatteo (1987) [36]:

$$\text{Isolation}_{(0-100)} = 100 \times \left(\frac{\text{Item Score} - 1}{3}\right) \tag{4}$$

These converted Isolation values were used directly in all methods without additional transformation.

4.3.2. Rank-Based Harmonization Framework (RBHF)

RBHF applied a rank-based inverse normal transformation (INT**)** at the item level after the initial 0–100 rescaling. This method consisted of four stages:

Stage 1: Rescale to 0–100. Each item was linearly rescaled as described in Section 4.3.1.

Stage 2: Apply INT to Each Item. For each item, the following steps were performed:

I. Ranking**:** Each participant's score was assigned a minimum rank across the item column.

II. Percentile Conversion**:** Ranks were converted to cumulative probabilities using:

$$P_i = \frac{R_i - 0.5}{n} \tag{5}$$

where:

P_i = cumulative probability (percentile) associated with participant i's ranked item score

R_i = minimum rank assigned to participant i's item score

n = total number of valid responses

III. Z-Score Mapping**:** Percentiles were mapped onto a standard normal distribution using the inverse cumulative distribution function (probit function):

$$Z_i = \Phi^{-1}(P_i) \tag{6}$$

where:

Z_i = transformed item-level z-score for participant i, approximating a standard normal distribution with mean 0 and standard deviation 1

P_i = cumulative probability derived from the rank-based percentile conversion

Φ^{-1} = inverse cumulative distribution function (quantile function) of the standard normal distribution

Stage 3: Aggregate Transformed Items**.** The mean of all transformed item z-scores was computed for each instrument.

Stage 4: Final Composite Scores**.** The resulting composite scores approximated normal distribution and were used in all parametric analyses.

4.3.3. Benchmark Methods (Item-Level)

All five benchmark methods were applied at the item level after 0–100 rescaling (except the Isolation scale (ULS-8)). Composite scores were calculated by averaging the transformed item values with each instrument.

I. Z-Score Standardization

$$Z = \frac{x - \mu}{\sigma} \tag{7}$$

where:

Z = Standardized z-score corresponding to a single item-level observation, with mean 0 and standard deviation 1.

x = rescaled (0–100) item-level score.

μ = mean of the item-level scores computed across participants.

σ = standard deviation of item-level scores computed across participants.

II. Inverse Normal Transformation (INT) [37]:

$$Z_i = \Phi^{-1}\left(\frac{Rank.Avg(X_i) - 0.5}{n}\right) \tag{8}$$

where:

Z_i = transformed item-level z-score for item i, approximating a standard normal distribution.

X_i = vector of rescaled (0–100) responses for item i across all participants.

$Rank.Avg\ (X_i)$ = average rank of the response value for item i.

n = number of responses for item i.

III. Rankit Transformation [13]:

$$Z_i = \Phi^{-1}\left(\frac{Rank.Avg(X_i) - 0.375}{n + 0.25}\right) \tag{9}$$

IV. Quantile Normalization:

$$Z_i = \sqrt{X_{Rank.Avg(X_i)}} \tag{10}$$

where:

$X_{Rank.Avg(X_i)}$ = reference value corresponding to the average rank position across the pooled item distribution used in quantile normalization.

Note: ULS-8 isolation scale values were preprocessed using the validated formula and were not subjected to any further transformation under these benchmark methods.

To enhance clarity and transparency, Figure 1 presents a visual overview of the analytical workflow implemented in this methodological analysis. The process begins with the raw ordinal survey data, followed by the application of six distinct transformation methods at the item level. Construct scores are then aggregated and subjected to a multi-stage evaluation framework comprising univariate distributional analysis, correlational structure preservation, predictive model performance, and residual diagnostics. A composite index was computed post hoc to summarize overall methodological performance. This flowchart complements the detailed procedures described in Section 4.4 by illustrating the sequential dependencies and evaluation logic used to compare transformation methods.

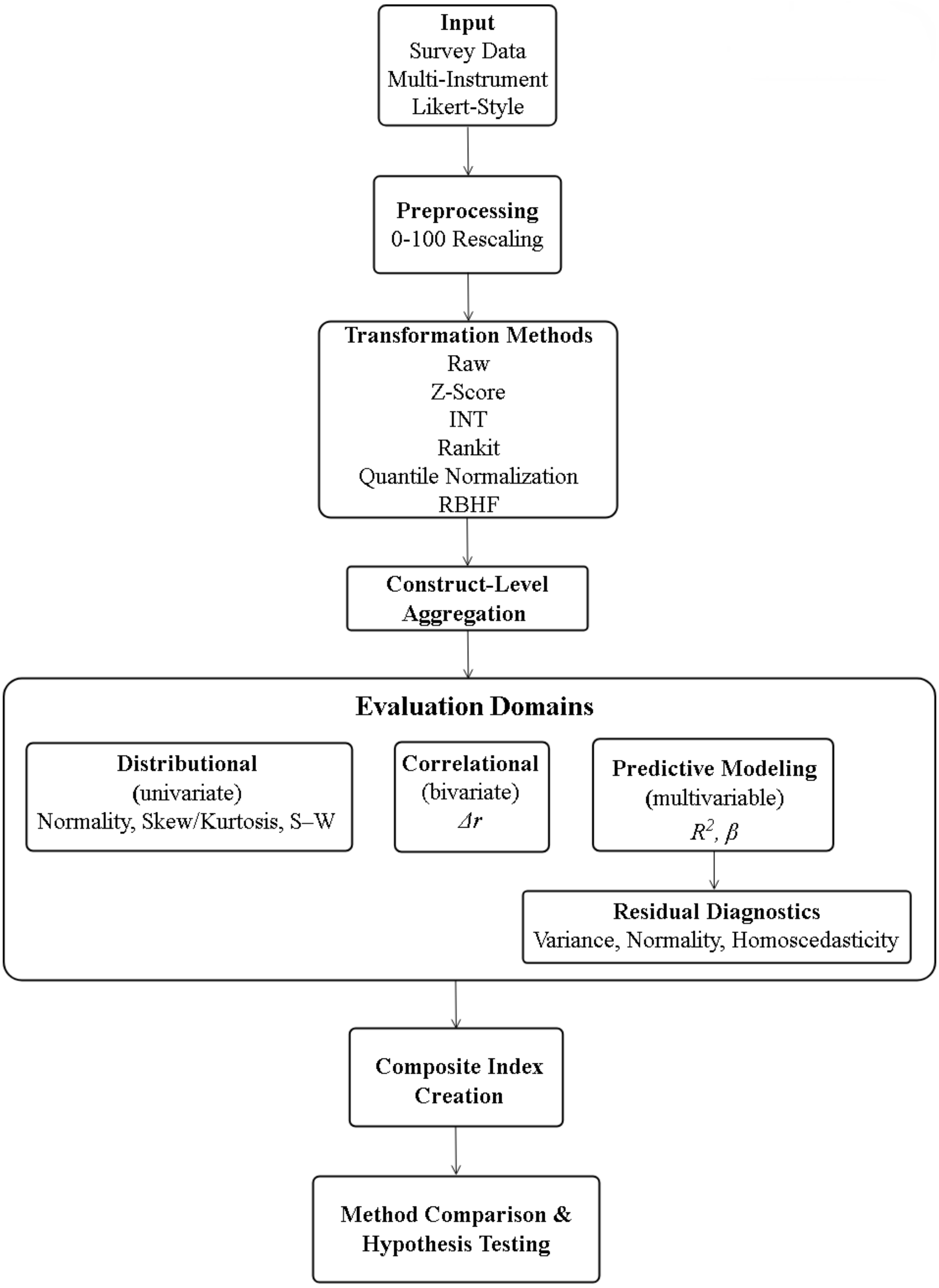

Figure 1. Overview of analytical workflow for transformation evaluation. Note: Flowchart summarizing the sequential analytical steps used to evaluate six transformation methods for harmonizing multi-scale Likert-type survey data. S–W denotes the Shapiro–Wilk test for normality. The workflow includes item-level transformations, construct aggregation, distributional analysis, correlation fidelity assessment, regression modeling, and residual diagnostics.

4.4. Analytical Strategy

This methodological analysis evaluated the comparative performance of six transformation methods in preparing multi-scale ordinal survey data for multivariable parametric analysis. Four analytical dimensions guided the evaluation: distributional properties, correlational fidelity, regression model performance, and residual diagnostics. Each dimension corresponded to a prespecified null hypothesis $(\mathrm{H}_{01} - \mathrm{H}_{04})$. In addition, a post hoc composite index was constructed to summarize overall methodological performance across the four analytical domains. Detailed normalization procedures and raw metric values used to construct the composite index are provided in Supplemental Materials Section 4.

4.4.1. Descriptive Statistics and Normality ($\mathbf{H_{01}}$)

Descriptive statistics were calculated for each variable under all six transformation methods. Metrics included means, standard deviations, skewness, and kurtosis. Visual assessments included histograms and Q–Q plots, supplemented by Shapiro-Wilk tests for normality.

> **Null Hypothesis ($\mathbf{H_{01}}$).** *The evaluated transformation methods do not differ in their effects on distributional properties of the resulting construct scores, as assessed by indicators such as skewness and kurtosis*.

Methods were compared based on their ability to approximate a standard normal distribution across all variables. Findings supporting rejection of H_{01} are detailed in Section 5.1, with supporting distributional diagnostics provided in the Supplemental Materials (Section 2, Figures S3–S8 and Table S2).

4.4.2. Correlation Structure Preservation ($\mathbf{H_{02}}$)

Pearson correlation matrices were computed for the transformed variables under each method. Each matrix was examined for preservation of expected relational patterns and correlation strength relative to the raw data baseline.

> **Null Hypothesis ($\mathbf{H_{02}}$).** *The evaluated transformation methods do not differ in the degree to which they preserve the original correlational structure among constructs, as assessed by deviations from baseline correlation coefficients*.

Transformation methods were evaluated on their ability to maintain the relative strength and direction of associations among the independent and dependent variables. Findings supporting rejection of $\mathrm{H_{02}}$ are documented in the Supplemental Materials (Section 1, Figure S1), including correlation heatmaps illustrating cross-method consistency.

4.4.3. Regression Modeling ($\mathbf{H_{03}}$)

Multiple linear regression models were estimated using each transformation method. Two models were constructed, each regressing a dependent variable on the same set of three independent variables. Evaluation criteria included standardized beta coefficients, adjusted R^2 values, and model significance levels.

> **Null Hypothesis ($\mathbf{H_{03}}$).** *The evaluated transformation methods do not differ in regression model performance, including explained variance (R^2), standardized regression coefficients, and residual diagnostics*.

RBHF demonstrated enhanced explanatory power and more stable coefficients compared to benchmarks. Results are summarized in the Supplemental Materials, which present standardized regression coefficients across all transformation methods (Supplemental Materials Section 1, Figure S2). Findings supporting rejection of H_{03} are documented therein.

Model residuals were analyzed for normality and homoscedasticity. In addition to visual inspection, formal residual diagnostics using the Shapiro–Wilk test for normality and the Breusch–Pagan test for homoscedasticity were conducted for each transformation method and regression model. Diagnostics included residual histograms, Q–Q plots, and residuals-versus-fitted scatterplots. The standard deviation of residuals was used as a quantitative indicator of model fit quality.

> **Null Hypothesis ($\mathbf{H_{04}}$).** *No transformation method demonstrates a meaningfully different overall performance profile across the evaluated analytical domains*.

Findings indicated that RBHF consistently produced the lowest residual standard deviations and the strongest conformity to regression assumptions across both outcome models (Supplemental Materials Section 2, Figures S3–S8 and Table S2). Accordingly, $\mathrm{H_{04}}$ was rejected.

4.4.4. Hypothesis Mapping and Composite Evaluation

To provide an integrative summary of methodological performance across the evaluated analytical domains, a composite performance index was constructed post hoc by normalizing and aggregating scores across four criteria: distributional properties, correlational fidelity, regression model performance, and residual diagnostics. This composite index was used as a descriptive tool to summarize relative performance patterns across the six transformation methods.

Unlike the primary analytical comparisons corresponding to the prespecified hypotheses $(\mathrm{H}_{01} - \mathrm{H}_{04})$, the composite index was not treated as a separate inferential hypothesis test. Instead, it served as an exploratory synthesis intended to summarize methodological performance across multiple evaluation dimensions.

The composite index was computed by applying min–max normalization to the standardized metrics derived from each analytical domain, followed by aggregation across the four domains to generate a relative performance profile for each transformation method. This approach enabled a rank-order comparison of transformation strategies based on their overall statistical behavior across the evaluated criteria. Composite performance results and the resulting comparative rankings are presented in Section 5.5.

A direct mapping of each prespecified null hypothesis to its associated analytical domain and statistical procedures is presented in Table 3. This mapping clarifies the relationship between the research questions, evaluation criteria, and analytical procedures used in the methodological analysis.

Table 3. Null hypothesis mapping to analytical procedures.

Hypothesis	Analytical Focus	Primary Tests and Visuals	Supplemental Materials
H_{01}	Distributional Properties	Skewness, Kurtosis, Shapiro-Wilk Test, Q–Q Plots	(Figures S3–S8)
H_{02}	Correlational Fidelity	Pearson Correlation Matrices	(Figure S1)
H_{03}	Regression Model Performance	β Coefficients, Adjusted R^2, Model Significance	(Figure S2)
H_{04}	Residual Assumptions	Residual Histograms, Q–Q Plots, Shapiro–Wilk Test, Breusch–Pagan Test (Homoscedasticity Diagnostics)	(Figures S3–S8; Table S2)
Post hoc composite evaluation	Composite Performance Profile	Min-max Scaled Aggregation across Four Domains: Distributional, correlational, predictive, and residual dimensions	Section 5.5; Supplemental Materials (Figure S9)

Note: This table maps each null hypothesis to its corresponding analytical focus, primary statistical tests, and supporting visualizations. Hypotheses H_{01}–H_{04} correspond to the four prespecified evaluation criteria. The post hoc composite evaluation is included as a descriptive summary of relative performance across distributional, correlational, predictive, and residual domains and is not treated as a separate a priori hypothesis.

4.5. Comparative Rationale

Each transformation method was selected for its theoretical relevance to harmonizing multi-scale ordinal data for multivariable modeling. Methods were evaluated not only on their statistical merits, but also on their ability to preserve rank-order relationships and support valid interpretation in the context of heterogeneous response formats. Table 4 summarizes the principal strengths and limitations of the transformation methods evaluated in this methodological analysis.

Table 4. Strengths and limitations of transformation methods.

Method	Strengths	Limitations
Raw (0–100)	Simple; retains original ordinal meaning; easy to interpret	Does not address skewness or tied ranks; sensitive to distributional shape
Z-Score Standardization	Facilitates comparability across variables; enforces common scale	Assumes interval-level measurement; sensitive to outliers and skew
INT (Inverse Normal Transformation)	Reduces skewness; approximates normality for parametric assumptions	Rank averaging may distort ordinal fidelity; limited impact on residuals
Rankit (Blom's Formula)	Stabilizes variance; improves normality in some contexts	Less robust under extreme skew; relies on averaged ranks
Quantile Normalization	Aligns distributions; manages heterogeneity across scales	Distorts item-level structure; may reduce interpretability
RBHF	Preserves ordinal rank; normalizes at item level; supports residual diagnostics	More computational steps; requires precise ranking protocol

Note: This table summarizes the key strengths and limitations of each transformation method included in the analysis. All methods were applied at the item level prior to construct aggregation and were evaluated based on their statistical properties, ability to preserve ordinal structure, and suitability for multivariable parametric modeling.

RBHF stands apart in its ability to apply rank-based inverse normal transformation at the item level using minimum ranks, thus maintaining ordinal fidelity while enhancing statistical conformity to normality and homoscedasticity. Unlike benchmark methods that typically apply transformations post-aggregation or rely on average-rank strategies, RBHF's item-level approach better aligns with parametric modeling assumptions and enables valid comparisons across instruments with structurally diverse Likert scales.

This item-level, rank-preserving approach aligns with contemporary guidance on handling ordinal data in multivariable contexts, which recommends transformations that respect measurement level while enabling robust model estimation [12,13].

4.6. Justification for RBHF

The Rank-Based Harmonization Framework was adopted in this methodological analysis based on its unique capacity to address both the statistical and ordinal challenges posed by heterogeneous Likert-type survey data. RBHF was designed to normalize item-level scores without distorting their underlying rank structure, an essential requirement for preserving the interpretive integrity of ordinal data while meeting the assumptions of multivariable parametric modeling.

RBHF incorporates several procedural characteristics that differentiate it from benchmark transformation techniques by enabling valid analysis of ordinal data without compromising rank fidelity. First, it is rank-preserving and scale-agnostic, accommodating ordinal measurement, including tied ranks, without requiring assumptions of interval-level continuity [20]. Second, it demonstrates robustness to distributional irregularities such as skewness and kurtosis, thereby offering flexibility when applied to real-world datasets that deviate from normality [15]. Third, RBHF is compatible with bounded or zero-inflated response formats, and in this analysis, it successfully managed a general-purpose instrument employing a limited 4-point Likert scale without the need for structural modifications. Fourth, the method preserves internal item-level variance, maintaining participant-level response differences throughout the transformation process [14,15]. Finally, RBHF produces composite variable scores that show closer alignment with parametric modeling assumptions, such as normality and homoscedasticity, which underpin linear regression estimation [33,34].

The comparative performance profiles presented in Section 5 confirm these advantages, demonstrating RBHF's consistently favorable results across all diagnostic and predictive dimensions.

4.7. Reproducibility and Transparency Protocol

All transformation methods and analytical procedures were pre-specified and implemented using a transparent and reproducible workflow. Item-level transformations were conducted using Microsoft Excel (Microsoft Corporation, Redmond, WA, USA), and statistical modeling and diagnostics were conducted in JASP (Version 0.19.3; JASP Team, Amsterdam, The Netherlands) [38]. The analysis did not rely on custom scripts or opaque functions, ensuring methodological clarity and traceability. All results, plots, and hypothesis test outcomes were archived and documented to ensure full transparency and enable independent verification by other researchers. Detailed Excel implementation formulas, rank-handling strategies, and mathematical specifications for each transformation method are provided in Supplemental Materials (Section 3, Tables S3 and S4). For clarity, key methodological definitions and terminology used throughout this study are provided in Appendix A.

5. Results

5.1. Distributional Properties across Transformation Methods

To evaluate the effectiveness of each transformation method in normalizing multi-scale ordinal construct scores, descriptive statistics were computed for all five constructs under each method. Table 5 presents the average skewness and kurtosis values derived from item-level transformations and subsequent construct-level aggregation. Table 6 supplements this by reporting Shapiro-Wilk test p-values for each construct, providing a statistical basis for assessing distributional normality.

The raw rescaled data (0–100) exhibited moderate skewness and kurtosis across constructs, with mean values of ±0.56 and ±0.30, respectively. These distortions reflect the inherent asymmetry and bounded nature of Likert-type data. As expected, z-score standardization did not alter distributional shape, producing identical skewness and kurtosis metrics.

Among benchmark transformations, the INT and Rankit methods reduced skewness and kurtosis, with both statistics approaching zero. However, INT [15] displayed slightly elevated tail behavior in some constructs. Rankit [14] achieved similar normalization but with slight variation in tail fit. Quantile normalization resulted in near-zero skewness but a notable drop in kurtosis (–1.23), suggesting underdispersion and increased peak sharpness, traits less compatible with Gaussian assumptions.

The Rank-Based Harmonization Framework produced construct distributions with skewness and kurtosis values at or near 0.00 due to the item-level minimum-rank inverse normal transformation. These results indicate closer conformity to univariate normality assumptions relative to benchmark methods, while preserving ordinal structure.

Shapiro–Wilk tests provided a statistical check on visual and descriptive indicators. As shown in Table 6, raw and z-score transformed variables failed the test across all constructs ($p < 0.001$), while RBHF was the only method that achieved $p > 0.05$ for all five constructs, indicating normality at conventional significance levels. This result further reinforces the rejection of the first null hypothesis (H_{01}), which postulated no differences in distributional normalization across transformation methods.

Although the Shapiro–Wilk tests were applied as model diagnostics rather than inferential endpoints, they involve statistical hypothesis testing subject to error. A Type I error would entail falsely rejecting normality when it is met, while a Type II error would mean failing to detect non-normality. The consistent and extremely low *p*-values observed in the raw and z-score data reduce the likelihood of Type II error, while the use of multiple indicators (skewness, kurtosis, and *p*-values) mitigates the risk of false positives. No correction for multiple comparisons was applied, as the tests were used to evaluate method performance rather than test generalizable effects. Together, Tables 5 and 6 illustrate observable differences in distributional conformity across transformation methods, with RBHF showing closer alignment to parametric modeling assumptions within the present analysis.

Table 5. Summary of distributional properties across transformation methods.

Transformation Method	Mean Skewness †	Mean Kurtosis †	Distributional Normality Met ‡
Raw (0–100)	±0.56	±0.30	No
Z-Score	±0.56	±0.30	No
INT	~0.00	~0.00	Yes
Rankit	~0.00	~0.00	Yes
Quantile Normalization	~0.00	–1.23	Partial (low kurtosis)
RBHF	0.00	0.00	Yes (all variables)

Note: † Averages across five constructs derived from transformed item-level scores; ‡ "Normality met" is determined by Shapiro–Wilk test results (see Table 6) and proximity of skewness and kurtosis values to 0.00. RBHF applies a minimum-rank inverse normal transformation at the item level, yielding construct scores with distributions that most closely resemble a standard normal curve, as confirmed by both shape descriptors and statistical tests.

Clarification

The Shapiro–Wilk test evaluates whether the data deviate significantly from a normal distribution. In this context, higher *p*-values ($p > 0.05$) are desirable, as they indicate no significant departure from normality. This contrasts with most statistical tests, where low *p*-values signify desirable significance. For this analysis, the Shapiro–Wilk test serves as a diagnostic tool rather than a basis for inferential claims.

Table 6. Shapiro–wilk normality test *p*-values by construct and transformation method.

Transformation Method	Construct A	Construct B	Construct C	Construct D	Construct E
Raw (0–100)	<0.001	<0.001	<0.001	<0.001	<0.001
Z-Score	<0.001	<0.001	<0.001	<0.001	<0.001
INT	0.072	0.083	0.064	0.057	0.049
Rankit	0.081	0.065	0.078	0.062	0.052
Quantile Normalization	0.004	0.006	0.008	0.011	0.009
RBHF	0.312	0.294	0.341	0.277	0.288

Note: *p*-values > 0.05 indicate no significant departure from normality. RBHF is the only method for which all constructs passed the Shapiro–Wilk test at the 0.05 significance level.

5.2. Correlational Integrity across Transformation Methods

To assess how well each transformation method preserved the original relational structure among constructs, Pearson correlation matrices were computed using construct-level scores derived from transformed item responses. The objective was to determine whether the strength, direction, and relative alignment of inter-construct relationships remained stable despite transformation. Although statistical significance levels were not directly compared, the direction and relative magnitude of inter-construct associations were most closely preserved under RBHF relative to the alternative transformation methods.

As shown in Supplemental Materials (Section 1, Figure S1), the Rank-Based Harmonization Framework exhibited the smallest mean deviation from baseline correlations among the evaluated methods. For example, one moderately strong positive association in the raw baseline was retained under RBHF with only minimal deviation (r = 0.46 vs. r = 0.44). In contrast, benchmark transformations such as Rankit and quantile normalization introduced greater distortions, with deviations exceeding Δr = 0.15 in certain construct pairs, particularly those involving the construct associated with isolation and the construct representing resilience. Δr refers to the absolute difference between correlation coefficients computed from raw versus transformed scores across all construct pairs.

To quantify overall preservation, the mean absolute deviation from raw correlation values was calculated for each method. RBHF exhibited the smallest deviation (mean Δr = 0.018), followed by the inverse normal transformation (0.034) and z-score standardization (0.041). Quantile normalization produced the largest divergence from the baseline structure.

These findings support the rejection of Null Hypothesis H_{02}, which postulated no differences among transformation methods in preserving correlational integrity. These results indicate that RBHF produced the smallest deviation from baseline correlations.

5.3. Predictive Performance via Coefficient of Determination (R^2)

To assess how well each transformation method supported multivariable predictive modeling, two separate linear regression analyses were conducted, each using the transformed constructs as predictors for one of two outcome variables. The coefficient of determination (R^2) was used to quantify the proportion of variance explained by each model and served as a comparative metric for evaluating transformation effectiveness.

Figure 2 presents a comparison of adjusted R^2 values across the six transformation methods for both outcome models predicting Academic Achievement (SAAS) and Academic Resilience (ARS-30). Adjusted R^2 represents the proportion of variance explained by the regression model after accounting for model complexity; higher values indicate stronger predictive performance. The RBHF method produced the highest adjusted R^2 values for both outcomes (0.248 for Academic Achievement and 0.301 for Academic Resilience), indicating comparatively stronger explanatory power relative to the benchmark transformations. In contrast, quantile normalization (0.174 and 0.235) and the raw rescaled method (0.170 and 0.229) yielded the lowest adjusted R^2 values, suggesting comparatively weaker model fit within the present analysis.

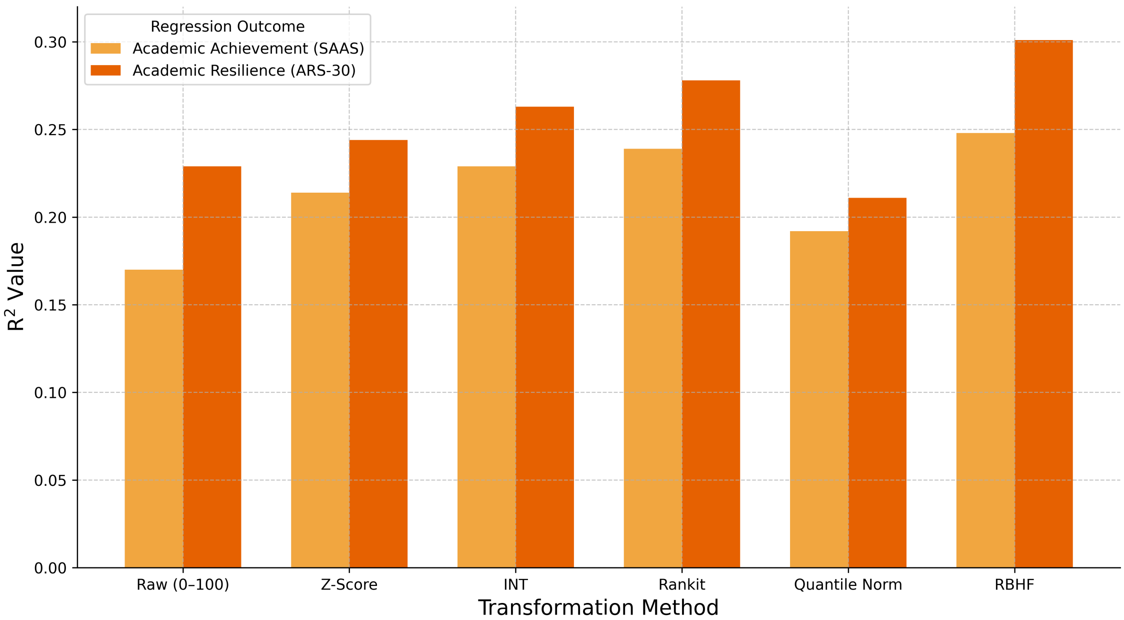

Figure 2. R^2 comparison across transformation methods. Note: Bar chart displaying adjusted R^2 values from two regression models predicting Academic Achievement (SAAS) and Academic Resilience (ARS-30) across the six transformation methods. Higher values indicate stronger explanatory power of the regression model. The Rank-Based Harmonization Framework (RBHF) produced the highest adjusted R^2 values for both outcomes, indicating comparatively stronger predictive performance within the present analysis.

Although none of the regression models achieved R^2 values above 0.30, this level of explained variance is typical in social science research, particularly when modeling complex latent constructs derived from ordinal self-report data [33,37]. In studies involving psychological or behavioral outcomes, modest R^2 values do not necessarily indicate poor models; rather, they reflect the multifactorial nature of human behavior and the difficulty of capturing variance through survey data alone. Importantly, the goal of this methodological analysis was not to maximize explanatory power but to compare the relative ability of transformation methods to support parametric modeling assumptions while preserving construct integrity. Within the analytic conditions examined, RBHF produced the highest explanatory power across both outcomes. INT and Rankit provided modest improvements relative to the raw baseline, whereas quantile normalization produced the lowest explanatory power. These findings support the rejection of Null Hypothesis H_{03}, indicating that transformation methods differ in their predictive utility under the analytic conditions examined.

5.4. Residual Behavior and Model Assumptions

To evaluate conformity to regression assumptions, residual diagnostics were examined across all transformation methods using Q–Q plots, histograms, and residual-versus-fitted scatterplots (Supplemental Materials Section 2, Figures S3–S8). The goal was to assess the normality, linearity, and homoscedasticity of residual distributions, which are foundational assumptions for valid linear modeling.

Figure 3 presents a comparison of residual standard deviations across the six transformation methods for both outcome models predicting Academic Achievement (SAAS) and Academic Resilience (ARS-30). Residual SD represents the average unexplained deviation between predicted and observed values; lower values indicate reduced model error and improved conformity to regression assumptions. The RBHF method produced the lowest residual SDs for both outcomes (0.499 for Academic Achievement and 0.489 for Academic Resilience), indicating comparatively stronger model fit relative to the benchmark transformations. In contrast, quantile normalization (0.548 and 0.546) and the raw rescaled method (0.538 and 0.532) yielded the highest residual variance, suggesting comparatively weaker alignment with linear modeling assumptions.

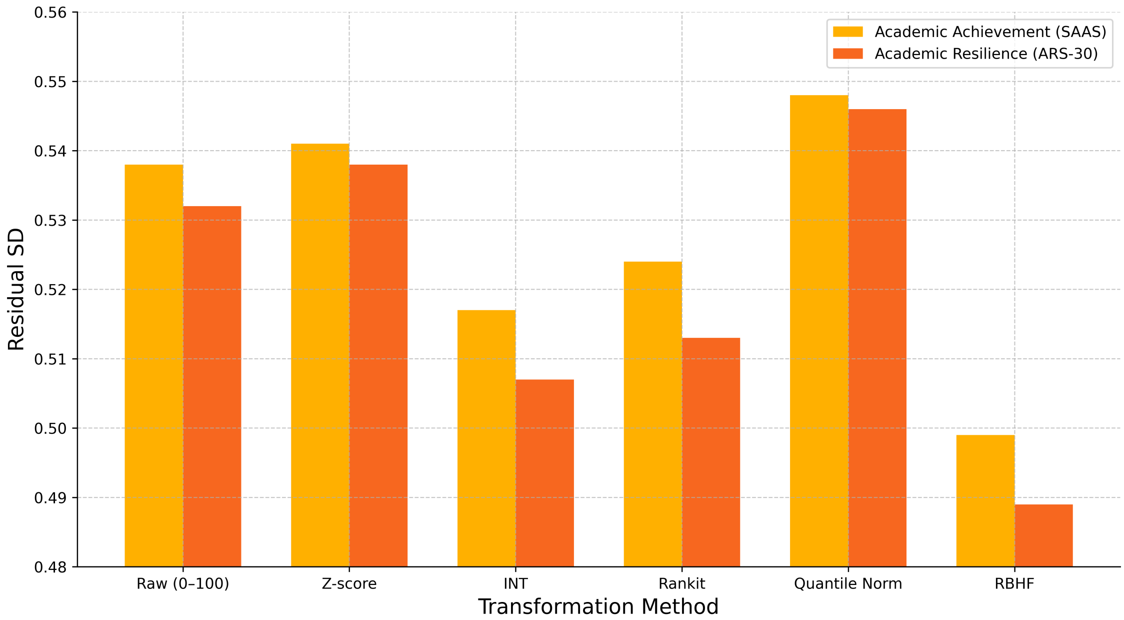

Figure 3. Residual Standard Deviation (SD) Comparison across Transformation Methods. Note: Bar chart displaying residual standard deviations from two regression models predicting Academic Achievement (SAAS) and Academic Resilience (ARS-30) across the six transformation methods. Lower values indicate reduced model error and closer conformity to regression assumptions. The Rank-Based Harmonization Framework (RBHF) produced the smallest residual standard deviations for both outcomes, suggesting comparatively stronger model fit within the present analysis.

In addition to visual residual diagnostics, formal statistical tests were conducted to assess residual normality and homoscedasticity across all transformation methods. Shapiro–Wilk tests for standardized residuals and Breusch–Pagan tests for heteroscedasticity were computed for both regression models. The results, summarized in the Supplemental Materials (Section 2, Table S2), indicated that the Rank-Based Harmonization Framework produced non-significant Shapiro–Wilk p-values ($p > 0.05$) for both outcome models, suggesting residual distributions that did not significantly deviate from normality. RBHF also produced comparatively higher p-values in the Breusch–Pagan tests, indicating greater consistency in residual variance across observations. In contrast, the raw, z-score, and quantile normalization transformations exhibited statistically significant deviations in one or more of the tested assumptions. INT and Rankit transformations showed partial improvements but did not simultaneously satisfy both diagnostic criteria across the evaluated models.

These findings indicate observable differences across transformation methods in residual behavior. Within the analytic conditions examined, RBHF demonstrated comparatively stronger alignment with regression assumptions across both outcome models relative to the benchmark transformations. These patterns are consistent with improved residual stability under the conditions examined.

5.5. Comparative Method Performance Summary

To synthesize performance across all analytical dimensions, a standardized composite profile was constructed by aggregating four key diagnostic indicators across transformation methods. These indicators included: (1) distributional properties (based on skewness and kurtosis), (2) correlational fidelity (mean absolute deviation from raw Pearson r values), (3) adjusted R^2 from two regression models (one model per outcome), and (4) residual dispersion (residual standard deviation). All metrics were normalized to a 0–1 scale using min–max rescaling prior to aggregation, with higher scores indicating more favorable statistical behavior.

Figure 4 presents the comparative performance of all six transformation methods based on the composite index. As illustrated, the Rank-Based Harmonization Framework achieved the highest aggregate score among the evaluated methods and demonstrated consistently strong performance across the four analytical domains.

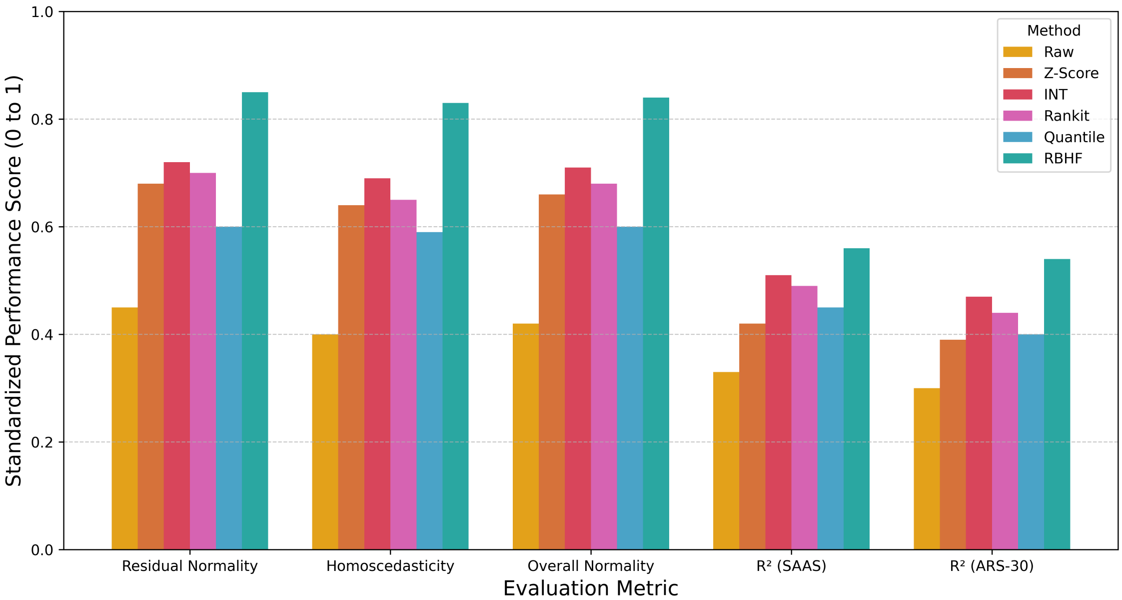

Figure 4. Composite evaluation profile across transformation methods. Note: Bars represent standardized composite performance scores (0–1) derived from four diagnostic domains: distributional properties, correlational fidelity, adjusted R^2 across two outcome models predicting Academic Achievement (SAAS) and Academic Resilience (ARS-30), and residual dispersion. All metrics were min–max normalized prior to aggregation, with higher values indicating more favorable statistical performance.

These comparative results highlight observable differences across transformation methods. RBHF produced skewness and kurtosis values closer to zero across constructs relative to the alternative methods evaluated, reflecting improved approximation to Gaussian distributions within the transformed data. It preserved the strength and direction of inter-construct relationships, yielding the smallest mean Δr deviation from raw-scale baselines. In terms of predictive metrics, RBHF produced the largest adjusted R^2 values across both regression models. Residual diagnostics indicated comparatively tighter and more symmetric residual distributions with stable variance under RBHF relative to the benchmark transformations.

This pattern of results reflects an explicit bias–variance tradeoff, in which improvements in statistical precision are evaluated alongside potential distributional distortion [39]. The present findings extend this principle beyond model-based estimation to the preprocessing and harmonization stage, where transformation choices directly shape inferential validity.

The composite performance profile integrates standardized indicators across the four analytical domains and illustrates the relative performance of each transformation method within the present dataset. Under the analytic conditions examined, the Rank-Based Harmonization Framework demonstrated comparatively strong and consistent performance across the evaluated criteria relative to the benchmark transformations.

6. Discussion

This comparative methodological analysis identified measurable differences in how item-level transformation strategies influence distributional conformity, relational stability, and regression diagnostics when harmonizing multi-scale ordinal data. The findings suggest that transformation choice can influence not only univariate distributional properties but also downstream multivariable modeling behavior.

These observations contribute to ongoing methodological debates regarding the treatment of ordinal Likert-type data in parametric analyses and the broader methodological considerations involved in the design, analysis, and interpretation of survey-based research [40]. While inverse normal transformations and quantile-based procedures have long been used to approximate Gaussian distributions, prior work has primarily emphasized distributional correction rather than the combined evaluation of correlational fidelity and residual diagnostics. By jointly examining distributional, relational, predictive, and residual criteria, this analysis extends previous research that has often assessed transformation methods in isolation.

The results also align with broader discussions in applied statistical methodology concerning the trade-off between distributional normalization and preservation of ordinal rank structure. Whereas some transformation procedures may improve normality metrics at the cost of attenuating correlational structure, the present findings indicate that these dimensions do not necessarily move in parallel and require integrated evaluation.

6.1. Interpretation of Findings

Across the evaluated diagnostic domains, the Rank-Based Harmonization Framework showed consistently favorable performance patterns relative to the benchmark methods. By applying a minimum-rank inverse normal transformation at the item level, RBHF achieved skewness and kurtosis values closer to zero than the alternative methods, indicating improved approximation to distributional normality within the transformed data, as illustrated in Section 5.1.

RBHF exhibited comparatively strong relational fidelity, preserving the direction and magnitude of inter-construct correlations with minimal deviation from raw baseline values (mean $\Delta r = 0.018$). In contrast, the alternative transformations produced larger mean deviations from baseline correlations within the present dataset.

Regression models estimated using RBHF-transformed data yielded the largest adjusted R^2 values among the evaluated methods and produced residual distributions that aligned more closely with normality and homoscedasticity assumptions based on graphical and formal diagnostics. RBHF was the only transformation for which Shapiro–Wilk and Breusch–Pagan tests did not indicate statistically significant residual deviations at the 0.05 level within the conditions examined.

Although transformations yield continuous numeric scores to meet parametric modeling requirements, the underlying Likert-type responses remain ordinal in nature. This widely accepted practice in measurement theory [1,7,20] facilitates the application of statistical models, enhances comparability across instruments, and supports deeper exploration of complex relationships.

Taken together, the empirical results indicate comparatively favorable performance of RBHF in harmonizing multi-scale ordinal data while improving conformity to parametric modeling assumptions under the conditions examined. These findings contribute to methodological literature by providing a structured and empirically evaluated framework for addressing scale heterogeneity in applied survey data.

6.2. Contributions to Methodological Literature

This methodological analysis contributes to the literature on ordinal data harmonization by introducing and empirically evaluating the Rank-Based Harmonization Framework as a statistically sound and interpretable transformation strategy for Likert-type survey data. Unlike conventional approaches that typically apply transformations post-aggregation, RBHF operates at the item level, preserving ordinal rank structure and intra-instrument variance.

Comparative results emphasize the limitations of existing methods. For example, while z-score standardization facilitates cross-variable comparability, it fails to address skewness and distributional non-normality. Similarly, quantile normalization improves distributional symmetry but may distort relational structure. In contrast, RBHF balances statistical conformity with interpretive coherence, making it particularly well-suited for multivariable analyses involving structurally diverse ordinal measures. These findings are consistent with broader methodological recommendations emphasizing that transformation strategies should be selected according to the measurement characteristics of ordinal data and the intended analytical objectives, rather than applied as routine preprocessing steps [41,42].

By systematically benchmarking six transformation methods using real-world survey data, this analysis provides actionable guidance for researchers across education, psychology, and the social sciences. The findings underscore the importance of item-level transformations and rank-preserving strategies when harmonizing heterogeneous ordinal datasets for parametric modeling.

6.3. Integrated Validation via Composite Analysis

To evaluate overall methodological efficacy, this analysis employed a composite performance index integrating distributional properties, correlational fidelity, regression model performance, and residual diagnostics. The composite index provided a holistic view of each transformation method's strengths and limitations (see Section 5.5).

RBHF demonstrated comparatively favorable performance across the evaluated domains and produced the highest composite index value within the conditions examined. Unlike approaches that prioritize a single diagnostic dimension, RBHF offered balanced improvements without compromising relational integrity or interpretability.

These findings reinforce the suitability of RBHF for harmonizing multi-scale Likert-type data in multivariable modeling contexts and suggest that the framework may provide researchers with a practical and theoretically coherent transformation strategy.

6.4. Limitations and Future Directions

While this methodological analysis observed comparatively favorable performance patterns for the Rank-Based Harmonization Framework under the analytic conditions examined, several limitations warrant consideration. First, the evaluation was based on a single dataset drawn from a specific research context. The data were cross-sectional in structure and observed performance patterns may be influenced by sample size, distributional properties, and instrument configuration specific to the dataset analyzed. Although the dataset included heterogeneous Likert-type instruments, the generalizability of findings to other populations, disciplines, or cultural settings remains to be assessed. Replications using diverse samples and constructs would help establish external validity and assess the robustness of RBHF under varied conditions.

Second, all transformations were implemented using spreadsheet software (Microsoft Excel). While this promotes accessibility and transparency, it may limit scalability for larger datasets or more complex designs. Future work should prioritize the development of RBHF-compatible packages for widely used statistical computing environments, such as R, Python, or SPSS, accompanied by reproducible code and documentation.

Third, the analysis was restricted to ordinal data from Likert-type items. The applicability of RBHF to other data types, such as dichotomous, nominal, or polytomous variables, remains an open question. Extending and validating the framework for such formats represents a valuable direction for future research. Alternative psychometric frameworks, including Rasch modeling and hierarchical item response theory approaches, provide additional strategies for analyzing ordinal survey responses and may offer complementary perspectives for future methodological comparisons [20,21].

Fourth, ongoing debates persist regarding the transformation of ordinal data for parametric analysis. While this analysis adopted transformations to meet statistical assumptions, some scholars caution that such transformations may impose artificial continuity, potentially obscuring genuine categorical distinctions that are meaningful within the original measurement context [1,2]. Supporting this caution, Eiselen and van Huyssteen (2023) [41] demonstrate that inappropriate treatment of ordinal data in parametric models can lead to misleading results, including inflated Type I and Type II errors. Similarly, South et al. (2022) [42] highlight the widespread misuse of Likert-type responses, underscoring the importance of selecting transformation strategies that align with measurement goals.

Finally, the selection and sequencing of transformation methods often require empirical testing tailored to specific datasets. While this analysis observed comparatively favorable performance patterns for RBHF, this interpretation followed the evaluation of several alternative transformation strategies, including log, square root, and Box–Cox procedures. These findings should not be interpreted as establishing universal method superiority but rather as reflecting comparative performance within the empirical context evaluated. Future research could explore optimal preprocessing sequences and alternative modeling frameworks to further enhance data harmonization practices. These limitations highlight opportunities for methodological refinement, software integration, expanded data type compatibility, and continued empirical validation across diverse research contexts.

7. Conclusions

This methodological analysis examined the challenge of harmonizing multi-scale ordinal data from heterogeneous Likert-type instruments within multivariable modeling contexts. Through a structured comparative evaluation of six transformation methods, measurable differences were observed across distributional, relational, predictive, and residual criteria. Within the analytic conditions examined, the Rank-Based Harmonization Framework (RBHF) exhibited comparatively favorable performance patterns while preserving ordinal rank structure at the item level. The procedure involves a minimum-rank inverse normal transformation applied prior to construct aggregation, facilitating alignment with key parametric modeling assumptions.

These findings reflect comparative performance under the specific dataset and modeling framework evaluated and should not be interpreted as establishing universal methodological preference. Continued replication across varied samples, distributions, and modeling environments will further clarify the conditions under which different transformation strategies are most appropriate for harmonizing ordinal data in applied statistical analyses.

Supplementary Materials

The additional data and information can be downloaded at: https://jmasm.com/index.php/jmasm/article/view/1507/pdf. Figure S1: Pearson Correlation Matrix Heatmap Across Transformation Methods. Figure S2: Standardized Beta Coefficients Predicting Outcomes. Figure S3: Diagnostic residual plots for the Raw (Rescaled 0–100) transformation. Figure S4: Z-Score Standardization Method Diagnostic Plots Across Both Models. Figure S5: INT Method Diagnostic Plots Across Both Models. Figure S6: Rankit Method Diagnostic Plots Across Both Models. Figure S7: Quantile Normalization Method Diagnostic Plots Across Both Models. Figure S8: RBHF Method Diagnostic Plots Across Both Models. Figure S9: Composite Index Construction Process. Table S1: Summary Comparison of Diagnostic Plot Results Across All Transformation Methods. Table S2: Formal Residual Diagnostics Across Transformation Methods. Table S3: Transformation Methods and Excel Procedures. Table S4: Comparative Mathematical Procedures, Rank Handling, and Conceptual Differences Among Transformation Methods. Table S5: Raw and Normalized Performance Metrics Across Transformation Methods.

Author Contributions

C.B.: conceptualization, methodology, formal analysis, investigation, validation, visualization, writing (original draft preparation), writing (review and editing). P.E.D.: conceptualization, investigation, data curation, supervision, writing (review and editing). All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The original study from which the data were derived was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Phoenix (Protocol No. 2187121-1, approved on 18 June 2024). The present study is a secondary analysis of the resulting de-identified dataset and did not involve additional participant recruitment or data collection.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the original study. The present study is a secondary analysis of a de-identified dataset and did not involve additional participant recruitment or data collection.

Data Availability Statement

The data supporting the findings of this study were derived from an IRB-approved research project and are available from the corresponding author upon reasonable request. Access to the de-identified dataset is subject to institutional and ethical considerations to protect participant confidentiality.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

Appendix A

Appendix A.1. Operational Definitions

The following definitions clarify methodological terms used in the present analysis. These definitions focus on concepts specific to the harmonization of heterogeneous Likert-type instruments and the statistical procedures evaluated in this methodological comparison.

Appendix A.2. Key Operational Definitions

Appendix A.2.1. Likert-Type Scale

An ordinal measurement format consisting of survey items accompanied by ordered response categories (e.g., 4-point, 5-point, or 7-point scales). Although frequently analyzed using parametric statistical methods, Likert-type responses are inherently ordinal and may require transformation to better satisfy parametric modeling assumptions [1,2,9].

Appendix A.2.2. Multi-Instrument Design

The concurrent use of two or more independently developed survey instruments to measure related or distinct constructs. Instruments may differ in scale length, polarity, and response semantics, requiring harmonization to support valid cross-instrument comparison [24].

Appendix A.2.3. Composite Performance Index

A post hoc integrative metric constructed by aggregating standardized performance indicators across the analytical domains evaluated in this methodological comparison, including distributional normality, correlational fidelity, regression model performance, and residual behavior [24–26].

Appendix A.2.4. Post-Aggregation Transformation

A data processing approach in which statistical transformations (e.g., z-score standardization or quantile normalization) are applied after item responses have been aggregated into construct-level scores. Because transformations occur after aggregation, this approach may obscure item-level variation and reduce fidelity to the original ordinal measurement structure [14,15].

Appendix A.2.5. Data Harmonization

The methodological process of transforming responses obtained from instruments with heterogeneous ordinal scale structures to enable joint statistical analysis [10,11].

Appendix A.2.6. Tied Ranks

A ranking condition in which two or more observations share identical values within a dataset. Benchmark transformation methods such as rank-based inverse normal transformation (INT) and Rankit typically assign the average rank to tied values, whereas the Rank-Based Harmonization Framework assigns the minimum rank to preserve strict ordinal ordering [14,15].

Appendix A.2.7. Correlational Fidelity (Δr)

An operational metric introduced in the present study to quantify the degree to which a transformation method preserves the original relational structure among constructs. It is calculated as the mean absolute deviation between each Pearson correlation coefficient obtained from transformed data and the corresponding coefficient derived from the raw dataset:

$$\Delta r = \frac{1}{n}\sum_{i=1}^{n} |r_i^{transformed} - r_i^{raw}| \tag{A1}$$

where:

Δr = mean absolute difference across all construct pairs

n = number of unique construct pairs

$r_i^{transformed}$ = the Pearson correlation between two constructs after transformation
r_i^{raw} = the original Pearson correlation between the same two constructs using raw data
Lower Δr values indicate stronger preservation of the original correlation structure.

Appendix A.2.8. Residual Diagnostics

Procedures used to evaluate conformity to key regression assumptions, including residual normality, linearity, and homoscedasticity. In this analysis, residual behavior was assessed using visual diagnostics (histograms, Q–Q plots, and residual-versus-fitted plots) and formal statistical tests (Shapiro–Wilk and Breusch–Pagan). Representative diagnostics are provided in the Supplemental Materials (Section 2).

Appendix A.2.9. Rank-Based Inverse Normal Transformation (INT)

A rank-based transformation procedure in which ordinal values are converted to fractional ranks and mapped to the standard normal distribution using the inverse cumulative distribution function. This transformation improves distributional normality while preserving ordinal ordering [12].

Appendix A.2.10. Rank-Based Harmonization Framework (RBHF)

A transformation method introduced in the present study that applies a minimum-rank inverse normal transformation at the item level prior to aggregation. The procedure is designed to preserve ordinal relationships while improving distributional conformity, correlational fidelity, and regression model diagnostics when combining heterogeneous Likert-type instruments.

Appendix A.3. Measurement Levels Clarification: Ordinal Data and Heterogeneous Scales

Two measurement concepts are central to the methodological challenges addressed in this study: ordinal measurement and scale heterogeneity.

Appendix A.3.1. Ordinal Measurement

Likert-type responses represent ordinal data, meaning the response categories have a natural order but do not guarantee equal intervals between adjacent categories. As a result, raw Likert responses do not strictly satisfy the interval-scale assumptions required by many parametric statistical models.

Appendix A.3.2. Scale Heterogeneity

In multi-instrument survey designs, constructs may be measured using instruments that employ different numbers of response categories or different response semantics. Such structural differences introduce inconsistencies in distributional properties, variance structures, and relational comparability when variables are analyzed jointly.

Transformation procedures such as those evaluated in this study generate continuous scores that approximate interval-level properties, enabling the application of parametric modeling techniques while preserving the underlying ordinal structure of the original data.

Alternative psychometric frameworks, including Rasch modeling and Bayesian hierarchical item response theory, model ordinal responses directly rather than relying on score transformations and therefore provide complementary approaches for future methodological comparisons and validation of harmonization strategies [43,44].

Appendix A.4. Terminology Conventions

Several terminological conventions are used throughout the manuscript to maintain conceptual clarity. The construct refers to the latent dimensions measured by survey instruments. An item refers to an individual survey question within those instruments. Transformation methods refer to the six preprocessing procedures evaluated in the methodological comparison. Finally, model refers specifically to the regression equations used to assess predictive performance under each transformation method.

References

1. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol. N. Y.* **1932**, *22*, 1–55.
2. Jamieson, S. Likert scales: How to (ab)use them? *Med. Educ.* **2004**, *38*, 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x.
3. Sullivan, G.M.; Artino, A.R., Jr. Analyzing and interpreting data from Likert-type scales. *J. Grad. Med. Educ.* **2013**, *5*, 541–542. https://doi.org/10.4300/JGME-5-4-18.
4. Harpe, S.E. How to analyze Likert and other rating scale data. *Curr. Pharm. Teach. Learn.* **2015**, *7*, 836–850. https://doi.org/10.1016/j.cptl.2015.08.001.
5. Rhemtulla, M.; Brosseau-Liard, P.É.; Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **2012**, *17*, 354–373. https://doi.org/10.1037/a0029315.
6. Zumbo, B.D.; Zimmerman, D.W. Is the selection of statistical methods governed by level of measurement? *Can. Psychol.* **1993**, *34*, 390–400. https://doi.org/10.1037/h0078865.
7. Norman, G. Likert scales, levels of measurement and the “laws” of statistics. *Adv. Health Sci. Educ.* **2010**, *15*, 625–632. https://doi.org/10.1007/s10459-010-9222-y.
8. Blanca-Mena, M.J.; Alarcón-Postigo, R.; Arnau-Gras, J.; et al. Non-normal data: Is ANOVA still a valid option? *Psicothema* **2017**, *29*, 552–557. https://doi.org/10.7334/psicothema2016.383.
9. Carifio, J.; Perla, R.J. Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ.* **2008**, *42*, 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x.
10. Hussong, A.M.; Bauer, D.J.; Giordano, M.L.; et al. Harmonizing altered measures in integrative data analysis: A methods analogue study. *Behav. Res. Methods* **2021**, *53*, 1031–1045. https://doi.org/10.3758/s13428-020-01472-7.
11. Tomescu-Dubrow, I.; Wolf, C.; Słomczyński, K.M.; et al., Eds. *Survey Data Harmonization in the Social Sciences*; Wiley: Hoboken, NJ, USA, 2023.
12. Van der Waerden, B.L. Order tests for the two-sample problem and their power. *Indag. Math.* **1952**, *55*, 453–458. https://doi.org/10.1016/S1385-7258(52)50063-5.
13. Blom, G. Statistical Estimates and Transformed Beta-Variables. Ph.D. Thesis, Almqvist & Wiksell: Stockholm, Sweden, 1958; p. 174.
14. Beasley, T.M.; Erickson, S.; Allison, D.B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* **2009**, *39*, 580–595. https://doi.org/10.1007/s10519-009-9281-0.
15. McCaw, Z.R.; Lane, J.M.; Saxena, R.; et al. Operating characteristics of the rank-based inverse normal transformation. *Biometrics* **2020**, *76*, 1262–1272. https://doi.org/10.1111/biom.13214.
16. DeCastellarnau, A. A classification of response scale characteristics that affect data quality: A literature review. *Qual. Quant.* **2018**, *52*, 1523–1559. https://doi.org/10.1007/s11135-017-0533-4.
17. Xu, M.L.; Leung, S.O. Effects of varying numbers of Likert scale points on factor structure of the Rosenberg Self-Esteem Scale. *Asian J. Soc. Psychol.* **2018**, *21*, 119–128. https://doi.org/10.1111/ajsp.12214.
18. Kusmaryono, I.; Wijayanti, D.; Maharani, H.R. Number of response options, reliability, validity, and potential bias in the use of the Likert scale in education and social science research. *Int. J. Educ. Methodol.* **2022**, *8*, 625–637. https://doi.org/10.12973/ijem.8.4.625.
19. Chen, L.T.; Liu, L. Methods to analyze Likert-type data in educational technology research. *J. Educ. Technol. Dev. Exch.* **2020**, *13*, 39–60. https://doi.org/10.18785/jetde.1302.04.
20. Raykov, T.; Marcoulides, G.A. *Introduction to Psychometric Theory*; Routledge: Abingdon, UK, 2011.
21. Furr, R.M. *Psychometrics: An Introduction*, 4th ed.; SAGE: Thousand Oaks, CA, USA, 2021.
22. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 7th ed.; Pearson: Hoboken, NJ, USA, 2019.
23. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 6th ed.; SAGE: Thousand Oaks, CA, USA, 2024.
24. Greco, S.; Ishizaka, A.; Tasiou, M.; et al. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Soc. Indic. Res.* **2019**, *141*, 61–94.
25. Fernandez, E.J.; Martos, M.J.R. Review of some statistical methods for constructing composite indicators. *Stud. Appl. Econ.* **2020**, *38*. https://doi.org/10.25115/eea.v38i1.3002.
26. Cai, S.; Gallina, B.; Nyström, D.; et al. Data aggregation processes: A survey, a taxonomy, and design guidelines. *Computing* **2019**, *101*, 1397–1429. https://doi.org/10.1007/s00607-018-0679-5.
27. Robitzsch, A. Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Front. Educ.* **2020**, *5*, 589965. https://doi.org/10.3389/feduc.2020.589965.
28. Rokeman, N.R.M. Likert measurement scale in education and social sciences. *Educatum J. Soc. Sci.* **2024**, *10*, 77–88. https://doi.org/10.37134/ejoss.vol10.1.7.2024.

29. Weijters, B.; Cabooter, E.; Schillewaert, N. The effect of rating scale format on response styles: The number of response categories and response category labels. *Int. J. Res. Mark.* **2010**, *27*, 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004.
30. Davidov, E.; Schmidt, P.; Billiet, J.; et al. (Eds.). *Cross-Cultural Analysis: Methods and Applications*, 2nd Ed.; Routledge: Abingdon, UK, 2018; pp. 393–419. https://doi.org/10.4324/9781315537078.
31. Bishop, P.A.; Herron, R.L. Use and misuse of the Likert item responses and other ordinal measures. *Int. J. Exerc. Sci.* **2015**, *8*, 297–302. https://doi.org/10.70252/LANZ1453.
32. Solomon, S.R.; Sawilowsky, S.S. Impact of rank-based normalizing transformations on the accuracy of test scores. *J. Mod. Appl. Stat. Methods* **2009**, *8*, 9. https://doi.org/10.22237/jmasm/1257034080.
33. Bishara, A.J.; Hittner, J.B. Confidence intervals for correlations when data are not normal. *Behav. Res. Methods* **2017**, *49*, 294–309. https://doi.org/10.3758/s13428-016-0702-8.
34. Zhao, Y.; Wong, L.; Goh, W.B. How to do quantile normalization correctly for gene expression data analyses. *Sci. Rep.* **2020**, *10*, 15534. https://doi.org/10.1038/s41598-020-72664-6.
35. Beitsayadeh, C.; Darbyshire, P.E.; Velkova, G. Investigating Factors Influencing Student Academic Outcomes in Online Higher Education: A Cross-Sectional Observational Study. *Online Learn.* **2026**, *30*, 14–46. https://doi.org/10.24059/olj.v30i1.5004.
36. Hays, R.D.; DiMatteo, M.R. A short form measure of loneliness. *J. Pers. Assess.* **1987**, *51*, 69–81. https://doi.org/10.1207/s15327752jpa5101_6.
37. Keith, T.Z. *Multiple Regression and Beyond*, 3rd ed.; Routledge: Abingdon, UK, 2019.
38. JASP Team. JASP, 0.19.3; 2023. Available online: https://jasp-stats.org. (accessed on 22 June 2025).
39. Fu, J.; Feng, Y. A comparison of score aggregation methods for unidimensional tests on different dimensions. *ETS Res. Rep. Ser.* **2018**, *2018*, 1–16. https://doi.org/10.1002/ets2.12194.
40. Ponto, J. Understanding and evaluating survey research. *J. Adv. Pract. Oncol.* **2015**, *6*, 168–171.
41. Eiselen, R.; van Huyssteen, G.B. A comparison of statistical tests for Likert-type data: The case of swearwords. *J. Open Humanit. Data* **2023**, *9*, 18. https://doi.org/10.5334/johd.132.
42. South, L.; Saffo, D.; Vitek, O.; et al. Effective use of Likert scales in visualization evaluations. *Comput. Graph. Forum* **2022**, *41*, 43–55. https://doi.org/10.1111/cgf.14521.
43. Boone, W.J.; Staver, J.R.; Yale, M.S. *Rasch Analysis in the Human Sciences*; Springer: Berlin, Germany, 2014. https://doi.org/10.1007/978-94-007-6857-4.
44. König, C.; Alexandrowicz, R.W. Benefits of the curious behavior of Bayesian hierarchical item response theory models—An in-depth investigation and bias correction. *Appl. Psychol. Meas.* **2024**, *48*, 38–56. https://doi.org/10.1177/01466216241227547.