



THE NEUTRAL OPTION AND COMPOSITE AGGREGATION OF LIKERT-TYPE RESPONSES: IMPLICATIONS FOR INFERENCE STATISTICAL ANALYSIS IN SOCIAL SCIENCE RESEARCH

Bashir M. Maalim, George M. Wanjobi

Lecturer, School of Business and Economics, Garissa University, Kenya

ABSTRACT

Likert-type scales are extensively used in academic research to measure attitudes, perceptions, behaviours, and other latent constructs; however, methodological decisions concerning the neutral response option and aggregation of multiple items remain insufficiently addressed in inferential statistical analysis. This analytical paper examines the role of the neutral option and composite aggregation of Likert-type responses in strengthening the validity, interpretability, and theoretical alignment of inferential analysis. Using a thematic content analysis and synthetic review of more than 20 recent journal articles and established statistics and research-methodology textbooks, the paper examines response-category design, neutral responses, measurement error, construct validity, reliability, composite scoring, multicollinearity, and regression modelling. The synthesis indicates that a neutral category can preserve meaningful respondent information where neutrality represents a legitimate position, although it should be distinguished from uncertainty, lack of knowledge, and non-applicability. The analysis further establishes that multiple statements measuring a theoretically coherent construct should, after appropriate reliability, dimensionality, and validity assessment, generally be aggregated into a composite score for construct-level inferential analysis. Composite aggregation improves parsimony, reduces unnecessary item-level redundancy, strengthens interpretability, and facilitates alignment between theoretical constructs and statistical models. Nevertheless, aggregation should not be applied mechanically because item-level predictors may provide advantages in prediction-oriented research. The paper concludes that valid neutral responses preserve the authenticity of measurement, while valid composite aggregation preserves the integrity of construct-level inference. Researchers are therefore encouraged to make response-category and aggregation decisions on theoretical, psychometric, and analytical grounds rather than convenience.

KEYWORDS: *Likert-Type Scale; Neutral Response; Composite Aggregation; Inferential Statistics; Regression Analysis.*

1.0 INTRODUCTION

Likert-type scales have become a widely used instrument in academic research for capturing attitudes, perceptions, beliefs, behaviours, and other latent constructs in measurable form. Their simplicity, flexibility, and capacity to translate subjective judgements into structured quantitative observations make them particularly valuable across social science, education, management, health, and public-sector research. However, their usefulness in inferential analysis extends beyond assigning numerical values to response categories; it also depends on how responses are structured, interpreted, and transformed into variables for statistical modelling. Two methodological decisions are particularly consequential: whether a neutral response option should be included and whether responses to multiple statements measuring the same construct should be aggregated into a composite score. These decisions influence measurement quality, reliability, statistical relationships, multicollinearity, and the validity and interpretation of regression and other inferential results (Creswell & Creswell, 2023; DeVellis & Thorpe, 2021; Field, 2024; Hair et al., 2022; Gökdemir & Yılmaz, 2023). Recent evidence indicates that a midpoint can demonstrate favourable psychometric properties when respondents use it meaningfully, while the number and structure of response categories may affect reliability, validity, and response behaviour (Kankaraš & Capecchi, 2025; Kusmaryono et al., 2022).

The methodological significance becomes more pronounced when Likert-type responses are used in correlation, regression, mediation, moderation, or other inferential procedures. A questionnaire may contain several statements under each independent variable, yet the theoretical framework may conceptualise these statements as indicators of a single underlying construct. Treating every statement as a separate predictor can unnecessarily increase model complexity, introduce correlations and multicollinearity among conceptually related variables, complicate interpretation, and create a mismatch between the theoretical and statistical models. Conversely, when theoretically coherent indicators are appropriately assessed and aggregated into a composite score, the resulting variable can more closely represent the construct specified in the research framework and provide a parsimonious basis for inferential analysis (Field, 2024; Hair et al., 2022; Kline, 2023; Tabachnick & Fidell, 2019). Nevertheless, composite aggregation should not be applied mechanically, as recent methodological evidence indicates that item-level predictors may sometimes provide advantages over composite scores, particularly when the primary objective is prediction rather than explanation (McClure et al., 2024).



Despite the extensive application of Likert-type scales, methodological practices concerning both the neutral option and composite aggregation remain inconsistent. Some researchers remove the midpoint to compel respondents to express a directional position, often assuming that neutrality represents indecision, avoidance, or poor-quality data, whereas others retain it on the premise that respondents may genuinely hold neither a positive nor negative position. Evidence suggests that neutral categories can be meaningfully used by respondents and may demonstrate favourable psychometric characteristics, although some respondents may use the midpoint as an escape option (Kankaraš & Capecchi, 2025). Similarly, some researchers aggregate items solely because they appear under the same variable heading, without adequately establishing conceptual coherence, dimensionality, reliability, and validity. Such mechanical aggregation can be as problematic as failing to combine indicators that genuinely represent the same construct (DeVellis & Thorpe, 2021; Kline, 2023). The methodological gap therefore concerns not whether Likert-type responses can be subjected to inferential analysis, but how individual responses should be appropriately transformed into construct-level variables for valid statistical inference. This paper consequently adopts an analytical and synthetic perspective to examine the complementary roles of the neutral option and composite aggregation.

2.0 METHODOLOGY

The paper adopted an analytical literature-review design based on content analysis and thematic synthesis. The objective was not to calculate an aggregate effect size but to identify, compare, and synthesise methodological arguments concerning the neutral response option, Likert-type scale construction, composite aggregation, measurement error, and inferential statistical modelling. Thematic analysis was used to identify recurring conceptual patterns, methodological disagreements, and areas of convergence across the literature (Braun & Clarke, 2021; Creswell & Creswell, 2023).

The literature comprised more than 20 sources, combining recent peer-reviewed journal articles with established statistical, psychometric, and research-methodology textbooks. The journal literature was prioritised from 2021 onward, except where an older methodological source was retained because it represents an established or foundational contribution. The textbook component deliberately included major authorities such as Field, Hair et al., DeVellis and Thorpe, Kline, Pallant, Tabachnick and Fidell, Creswell and Creswell, Gravetter and Wallnau, Saunders et al., and Sekaran and Bougie. This combination was considered appropriate because journal literature provides contemporary empirical and methodological evidence, whereas major textbooks provide established principles for measurement, regression, multivariate analysis, scale development, and research design (DeVellis & Thorpe, 2021; Field, 2024; Hair et al., 2022).

The analysis followed five thematic stages. First, literature concerning the number and nature of response categories was examined, with particular attention to midpoint and neutral responses. Second, studies concerning reliability, validity, dimensionality, and scale construction were synthesised. Third, literature concerning composite and aggregated scores was examined. Fourth, the implications of aggregation for measurement error, multicollinearity, regression, and statistical power were analysed. Finally, the literature was integrated into a conceptual framework linking questionnaire design to inferential interpretation.

The analytical process was synthetic rather than merely descriptive. Sources with similar arguments were combined to identify methodological convergence, while contrasting positions were explicitly examined. For example, evidence supporting composite scores was considered alongside recent research showing that item-level predictors can sometimes outperform composites in predictive modelling. Similarly, arguments supporting neutral categories were examined alongside evidence that some respondents may use midpoint responses strategically. This approach allows the paper to move from the question of *what researchers do* towards the more consequential question of *when and why a particular methodological choice is defensible* (Kankaraš & Capecchi, 2025; McClure et al., 2024).

3.0 LITERATURE REVIEW

3.1 Likert-Type Scales as Measures of Latent Constructs

The Likert scale was introduced by Rensis Likert in 1932 through his seminal work where he developed the technique to provide a relatively simple and systematic method for measuring attitudes and opinions by asking respondents to indicate their degree of agreement or disagreement with a series of statements, thereby converting subjective attitudes into quantifiable scores (Likert, 1932). The approach offered important practical advantages over earlier attitude-scaling techniques, particularly because it was easier to construct, administer, and score and did not require the use of expert judges.

Despite its widespread adoption, the Likert approach has attracted methodological criticism. A fundamental criticism concerns the ordinal nature of individual Likert-type responses: categories such as “strongly disagree” to “strongly agree” indicate ordered positions but do not necessarily establish equal distances between categories. Consequently, some scholars have questioned the appropriateness of calculating means or applying parametric procedures such as correlation and regression directly to individual Likert items (Boone & Boone, 2012; Jamieson, 2004). However, Norman (2010) challenges an overly restrictive interpretation of this criticism, arguing that parametric procedures can be robust when applied appropriately, particularly when multiple Likert items are combined into scales and sample sizes are adequate. Thus, the enduring methodological debate is not simply whether Likert-type data can be analysed statistically, but how individual ordinal responses should be measured, combined, and transformed into defensible variables for inferential analysis.



A fundamental premise of quantitative research is that theoretical constructs must be operationalised through observable indicators. Constructs such as leadership effectiveness, employee engagement, digital capacity, organisational culture, and service quality cannot normally be observed directly; researchers therefore develop multiple items intended to capture their conceptual domains. DeVellis and Thorpe (2021) argue that scale development requires careful specification of the construct, item generation, assessment of dimensionality, reliability, and validity before a scale is treated as a measure of the intended construct. Similarly, Creswell and Creswell (2023) locate operationalisation within the broader process of connecting research questions, variables, measurement, and statistical analysis.

This distinction between item and construct is central to the aggregation argument. A questionnaire statement is an observed indicator, whereas the construct is the theoretical concept that the collection of indicators is intended to represent. Hair et al. (2022) and Kline (2023) emphasise this distinction in multivariate and latent-variable analysis. Accordingly, the existence of six statements does not necessarily imply the existence of six explanatory variables; the six statements may constitute six observations of one underlying construct.

3.2 The Neutral Option and Preservation of Respondent Information

The neutral category should be understood as a measurement decision rather than a statistical nuisance. Where respondents can legitimately occupy an intermediate position, removing the midpoint forces them to select an alternative that may not represent their actual judgement. Kankaraš and Capecchi (2025) provide particularly strong recent evidence: their experimental comparison of scales with and without a neutral option found some improvement in reliability and factor-related psychometric characteristics when the midpoint was available, while most respondents appeared to use the category validly.

This position is supported by broader evidence concerning response-category design. Kusmaryono et al. (2022), after reviewing 60 studies, concluded that the number of response options has implications for reliability, validity, and potential response bias. Aybek and Toraman (2022), using item-response-theory analysis, found that five response categories provided advantages over three categories in reliability and information while not showing substantial disadvantages compared with seven categories. Thus, the evidence does not support a universal rule that more response categories or forced-choice formats are inherently superior.

Recent research also suggests that the meaning of response categories matters. Tutz (2021) notes that Likert-type categories are ordered and that the neutral middle category has a specific role that should be recognised in modelling. Similarly, Abulela and Khalaf (2024), in a scoping review, argue that the optimal number of response categories remains unresolved and should be considered in relation to validity evidence rather than determined through an arbitrary rule.

The appropriate conclusion is therefore conditional. A neutral option should be retained when neutrality is substantively meaningful, but it should not be used as a universal substitute for “don't know,” “not applicable,” or lack of information. Saunders et al. (2019) and Sekaran and Bougie (2020) emphasise the importance of designing response options that correspond to the information respondents can reasonably provide. Neutrality is consequently a valid response state only when the construct and wording make such a state meaningful.

3.3 Composite Aggregation of Multiple Indicators

Composite aggregation involves combining responses to several items intended to measure the same construct, typically through a sum or arithmetic mean. DeVellis and Thorpe (2021) treat composite scale construction as a central component of measurement development, while Hair et al. (2022) emphasise the role of multiple indicators in representing multidimensional and latent constructs. A composite therefore represents more than a mathematical transformation; it expresses the researcher's theoretical proposition that several observations collectively represent one construct.

The justification for aggregation should consequently be based on conceptual and empirical coherence. Items should first be linked theoretically to the same construct and then examined empirically through dimensionality, reliability, and validity procedures. McNeish (2023) stresses that scale validation is essential when using categorical or Likert-type indicators, while Kline (2023) demonstrates that factor structure provides evidence about the relationship between observed indicators and latent constructs.

The literature also cautions against treating internal consistency as sufficient evidence for aggregation. A high Cronbach's alpha does not prove that a scale measures the intended construct. DeVellis and Thorpe (2021) distinguish reliability from validity, while Hair et al. (2022) similarly emphasise construct validity and dimensionality. Consequently, researchers should not simply calculate alpha and average all items that happen to produce an acceptable coefficient.

3.4 Composite Aggregation and Measurement Error

The strongest statistical rationale for composite aggregation concerns measurement error. Individual questionnaire items are susceptible to wording effects, respondent interpretation, temporary circumstances, and item-specific variance. Glas et al. (2024) demonstrate that measurement error arising from items and raters can attenuate regression coefficients and reduce the power of linear models.

When several reliable indicators measure the same construct, aggregation can reduce the influence of item-specific noise and produce a more stable representation of the underlying attribute. This does not eliminate measurement error, but it can improve the signal-to-noise relationship of the observed construct. Hair et al. (2022), Field (2024), and DeVellis and Thorpe (2021) therefore provide complementary arguments for treating a validated composite as preferable to an isolated item when the research objective concerns a broad theoretical construct.

A more sophisticated alternative is latent-variable modelling, where measurement error can be explicitly represented rather than implicitly absorbed into an observed composite. Kline (2023) demonstrates this through structural equation modelling. Thus, composite aggregation can be viewed as a practical observed-variable solution, while latent-variable modelling offers a more explicit statistical treatment of measurement error where resources and research objectives permit.

3.5 Composite Aggregation and Regression Modelling

The argument for composite values becomes particularly compelling when regression hypotheses are formulated at the construct level. Suppose digital infrastructure is measured using six Likert-type statements. If the hypothesis states that “digital infrastructure influences service delivery,” entering all six statements separately into regression produces six predictors for one theoretical construct. The statistical model consequently becomes disconnected from the theoretical model.

A construct-level composite resolves this problem. If the six items are validated as indicators of digital infrastructure, their mean can be calculated as:

$$DI = \frac{DI_1 + DI_2 + DI_3 + DI_4 + DI_5 + DI_6}{6} \quad \text{Equation 1}$$

$$DC = \frac{DC_1 + DC_2 + DC_3 + DC_4 + DC_5 + DC_6}{6} \quad \text{Equation 2}$$

$$CE = \frac{CE_1 + CE_2 + CE_3 + CE_4 + CE_5 + CE_6}{6} \quad \text{Equation 3}$$

The regression model can then be specified as:

$$SD = \beta_0 + \beta_1 DI + \beta_2 DC + \beta_3 CE + \varepsilon \quad \text{Equation 4}$$

where (SD) represents service delivery, (DI) digital infrastructure, (DC) digital capacity, and (CE) citizen engagement. This model is more closely aligned with the conceptual framework because each coefficient corresponds to a theoretical construct rather than an isolated questionnaire statement (Field, 2024; Hair et al., 2022).

Composite aggregation also improves parsimony. If 20 questionnaire items represent four theoretical constructs, a four-predictor regression model is more interpretable than a 20-predictor model, assuming the aggregation is psychometrically justified. Tabachnick and Fidell (2019) and Hair et al. (2022) emphasise the importance of appropriate predictor specification, while Field (2024) demonstrates the consequences of excessive predictor intercorrelation for regression interpretation.

3.6 Aggregation and Multicollinearity

Items designed to measure the same construct are expected to correlate. Entering them simultaneously into regression can therefore generate substantial item-level multicollinearity. Highly correlated predictors can inflate standard errors, make regression coefficients unstable, and complicate the interpretation of unique effects (Field, 2024; Hair et al., 2022; Tabachnick & Fidell, 2019). Aggregation can reduce this unnecessary redundancy by transforming several correlated indicators into one construct-level predictor. However, aggregation should not be presented as a universal remedy for multicollinearity. Distinct constructs can remain highly correlated after aggregation; for example, digital infrastructure may correlate strongly with digital capacity. Researchers should therefore continue to examine correlation matrices, tolerance, variance inflation factors, and other relevant diagnostics after composite variables have been created (Field, 2024; Hair et al., 2022).

3.7 Composite Values and Inferential Interpretation

Composite aggregation also strengthens substantive interpretation. If five items measure employee engagement, a regression coefficient associated with their composite can be interpreted as the estimated relationship between employee engagement as a construct and the dependent variable. By contrast, a coefficient for one individual item describes only the relationship associated with that particular statement.

This distinction is particularly important in academic research where hypotheses derive from theories rather than individual questionnaire statements. Creswell and Creswell (2023) emphasise coherence between theoretical propositions, variables, and analytical methods, while Field (2024) and Hair et al. (2022) emphasise appropriate model specification. Composite aggregation therefore creates an analytical bridge between theoretical abstraction and statistical estimation.



The argument for composite scores should not be overstated. McClure et al. (2024) demonstrate mathematically that composite predictors can sometimes be inferior to item-level predictors in prediction-focused linear regression because aggregation can introduce prediction bias. Their conclusion is especially relevant where items are heterogeneous and the objective is prediction rather than explanation.

This finding introduces a crucial distinction between explanatory research and predictive research. Where the purpose is to test whether a theoretical construct influences another construct, validated composite measures are often conceptually appropriate. Where the objective is to maximise predictive accuracy, however, retaining item-level information or using alternative predictive techniques may sometimes be preferable. Kramer and Weldon (2023) likewise emphasise that constructing a composite requires careful consideration of which measures should be combined and how the composite should be constructed.

4.0 FINDINGS AND DISCUSSIONS OF THE REVIEW.

The thematic synthesis produced five principal findings. First, the neutral option should be understood as a potentially valid response state rather than automatically as a source of bias. Second, the neutral category should be distinguished from “don't know,” “not applicable,” and missing responses. Third, multiple Likert-type statements should be treated as indicators of constructs rather than automatically as independent variables. Fourth, validated composite aggregation can improve conceptual alignment, parsimony, interpretability, and potentially the stability of inferential estimates. Fifth, aggregation should remain conditional because item-level information can be valuable when the research objective is prediction or when indicators are conceptually heterogeneous (Kankaraš & Capecci, 2025; McClure et al., 2024).

The first finding challenges the common methodological assumption that a neutral response is necessarily undesirable. Recent experimental evidence indicates that many respondents use midpoint categories meaningfully and that their inclusion can improve some psychometric characteristics (Kankaraš & Capecci, 2025). This suggests that forcing respondents to choose between agreement and disagreement can sometimes create artificial variance rather than improve measurement. The implication for researchers is not that every Likert scale must contain a midpoint, but that midpoint removal requires substantive justification.

The second finding concerns information preservation. A legitimate neutral response should ordinarily be coded as a valid observation rather than automatically converted to missing. Tutz (2021) specifically identifies the neutral middle category as an important feature of Likert-type response structures, while Saunders et al. (2019) emphasise matching response alternatives to the information respondents can provide. Treating valid neutrality as missing changes the observed distribution and may systematically exclude respondents whose true position is intermediate.

The third finding concerns the distinction between item-level information and construct-level information. The fact that a questionnaire contains several statements does not mean that each statement constitutes a separate theoretical variable. DeVellis and Thorpe (2021), Hair et al. (2022), and Kline (2023) collectively support the proposition that multiple observed indicators can represent a common construct when theory and empirical evidence support such a structure. This distinction is particularly important for regression because the unit of statistical analysis should correspond as closely as possible to the unit of theorisation.

The fourth finding is that composite aggregation can provide a stronger basis for construct-level inferential analysis. A mean or sum of validated indicators reduces the number of predictors, limits unnecessary item-level redundancy, and makes regression coefficients easier to interpret. Measurement-error research further suggests that rating-scale error can attenuate regression coefficients, providing a statistical rationale for improving measurement quality before modelling (Glas et al., 2024).

The fifth finding introduces an important boundary condition. Composite aggregation is not synonymous with statistical superiority. McClure et al. (2024) show that composite predictors may sometimes perform worse than item-level predictors when the objective is prediction. The appropriate choice therefore depends on the purpose of the research. Construct-oriented explanatory research generally provides a stronger rationale for validated composites, whereas prediction-oriented research may justify retaining item-level information or comparing alternative representations of the predictors.

Taken together, these findings support a measurement-to-inference principle as shown in figure 1 below:

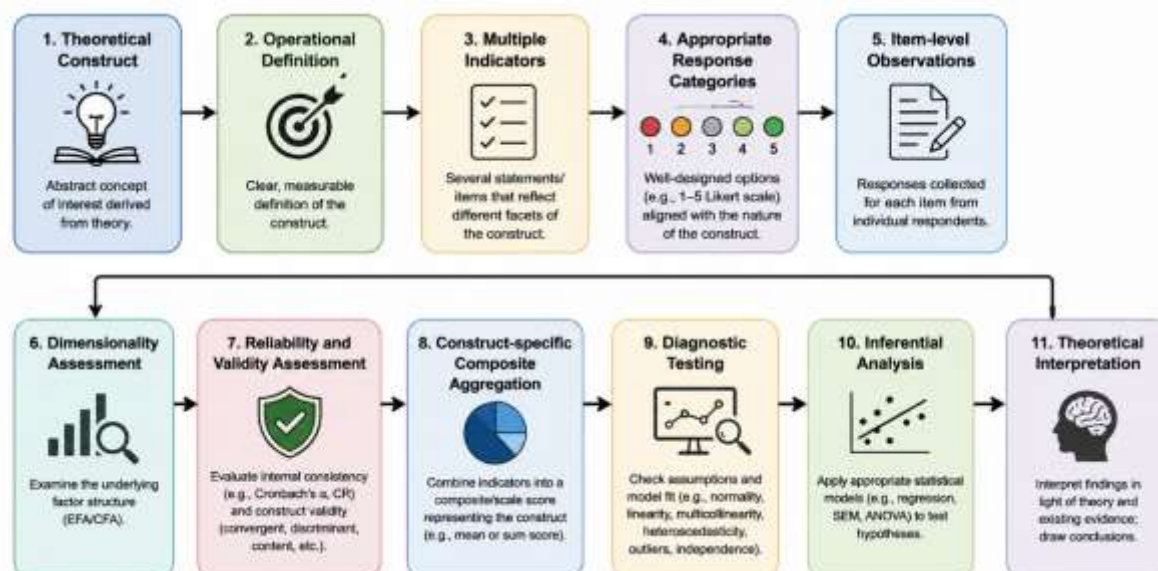


Figure 1
Measurement-to-Inference Principle

This sequence is preferable to the common but methodologically weaker approach of coding questionnaire responses and immediately entering every item into regression. It also clarifies that aggregation is not an arbitrary data-reduction procedure. It is a theoretically informed transformation that should occur only after evidence demonstrates that the indicators represent the same construct (DeVellis & Thorpe, 2021; Hair et al., 2022; Kline, 2023).

The findings also suggest that the mean is particularly useful as a composite value where all indicators employ the same Likert metric. For example, if six statements are scored from 1 to 5, the mean remains on the 1–5 metric and can therefore be interpreted using the same conceptual response continuum. The statistical difference between a mean and a sum is generally a linear transformation when the same items are included, so the methodological priority should be the validity of aggregation rather than an overly rigid preference for either calculation (Field, 2024; Pallant, 2020).

An additional implication concerns the treatment of neutral responses within composite scores. A respondent may select “neutral” on one item while selecting agreement on four others. It would therefore be inappropriate to conclude from the single midpoint response that the respondent is neutral regarding the entire construct. Composite aggregation allows the respondent's position across all relevant indicators to be represented. The neutral option and composite score consequently operate at different analytical levels: the former preserves item-level authenticity, while the latter summarises construct-level position (DeVellis & Thorpe, 2021; Kankaraš & Capecchi, 2025).

The analysis further suggests that researchers should distinguish between descriptive and inferential uses of Likert data. Item-level frequencies, percentages, means, and standard deviations can provide valuable descriptive information about how respondents reacted to individual statements. However, when hypotheses concern broader constructs, inferential tests should generally employ validated construct-level scores. This approach allows researchers to retain item-level information for descriptive interpretation while using construct-level variables for regression, correlation, mediation, moderation, or related analyses (Field, 2024; Hair et al., 2022).

Finally, composite aggregation should be embedded within a broader diagnostic framework. A researcher should establish content validity, dimensionality, reliability, and construct validity before aggregation; examine missing data and response patterns; assess correlations and multicollinearity after aggregation; and evaluate the assumptions appropriate to the chosen inferential model. Aggregation therefore improves the analytical structure but does not eliminate the need for statistical diagnostics (Field, 2024; Pallant, 2020; Tabachnick & Fidell, 2019).

5.0 CONCLUSION

This analytical review concludes that the neutral option and composite aggregation should be regarded as complementary methodological components of high-quality Likert-type measurement rather than as competing statistical choices. The neutral category can preserve legitimate respondent information when neutrality represents a meaningful position on the measured construct. Contemporary empirical evidence does not support the blanket assumption that midpoint responses are inherently invalid or that forced-choice formats are universally superior (Kankaraš & Capecchi, 2025; Kusmaryono et al., 2022).



The review further concludes that the strongest rationale for composite aggregation arises when several statements are theoretically designed and empirically demonstrated to measure the same construct. In such circumstances, an aggregated mean or sum provides a more theoretically meaningful predictor than treating each statement as an independent variable. Composite aggregation can reduce unnecessary predictor redundancy, improve parsimony, facilitate interpretation, and potentially mitigate the influence of item-specific measurement error (DeVellis & Thorpe, 2021; Field, 2024; Glas et al., 2024).

The overarching methodological proposition of this paper is therefore: A valid neutral response preserves the authenticity of measurement, while a valid composite score preserves the integrity of construct-level inference.

6.0 RECOMMENDATION

Researchers should retain the neutral option when genuine neutrality is theoretically possible, but distinguish it from “don't know,” “not applicable,” and missing responses. Researchers should not recode valid neutral responses as missing merely because they prefer directional responses. At the same time, researchers should avoid mechanical aggregation and should aggregate only those items that theory, dimensionality, reliability, and validity evidence indicate belong to the same construct (DeVellis & Thorpe, 2021; Hair et al., 2022; Kline, 2023).

For regression-based explanatory research, researchers are recommended to formulate hypotheses at the construct level and, where measurement evidence permits, calculate one composite value per construct rather than entering every statement as a separate predictor. This produces closer alignment among the conceptual framework, questionnaire, variables, and regression model. Nevertheless, researchers should retain item-level analysis where individual statements have independent theoretical meaning or where prediction rather than explanation is the primary objective (McClure et al., 2024).

Researchers should also report the aggregation process transparently, including the theoretical definition of each construct, number of indicators, response scale, treatment of reverse-coded items, evidence of dimensionality, reliability and validity, and the mathematical procedure used to obtain the composite value. Such reporting enables readers to evaluate whether the transformation from item responses to inferential variables is methodologically defensible.

7.0 SUGGESTIONS FOR FURTHER RESEARCH

The study suggests to include experimental studies comparing five-point Likert scales with and without neutral categories across different cultural, organisational, educational, and public-sector contexts; simulation studies examining how neutral-response proportions influence regression coefficients; comparative studies of item-level versus composite predictors under different sample sizes and levels of measurement error; and empirical studies comparing simple mean composites, weighted composites, factor scores, and latent-variable estimates. Further research should also investigate whether the value of the neutral category differs according to the sensitivity, complexity, or familiarity of the construct being measured. Such studies would extend the emerging evidence base and help establish context-sensitive guidelines for Likert-type instrument design and inferential analysis (Abulela & Khalaf, 2024; Kankaraš & Capecchi, 2025; McClure et al., 2024).

REFERENCES

1. Abulela, M. A. A., & Khalaf, M. A. (2024). Does the number of response categories impact validity evidence in self-report measures? A scoping review. *SAGE Open*, 14(1), 1–15. <https://doi.org/10.1177/21582440241230363>
2. Aybek, E. C., & Toraman, C. (2022). How many response categories are sufficient for Likert type scales? An empirical study based on the item response theory. *International Journal of Assessment Tools in Education*, 9(2), 534–547. <https://doi.org/10.21449/ijate.1132931>
3. Boone, H. N., Jr., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, 50(2), Article 48. <https://doi.org/10.34068/joe.50.02.48>
4. Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
5. Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
6. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.
7. Doğan, N., Yurtçu, M., & Gündeğer, C. (2023). The effect of option differences on psychometric properties of items in Likert-type scales. *Sakarya University Journal of Education*, 13(2), 207–237. <https://doi.org/10.19126/suje.1253876>
8. Field, A. (2024). *Discovering statistics using IBM SPSS Statistics* (6th ed.). SAGE Publications.
9. Glas, C. A. W., Jorgensen, T. D., & ten Hove, D. (2024). Reducing attenuation bias in regression analyses involving rating scale data via psychometric modeling. *Psychometrika*, 89, 42–63. <https://doi.org/10.1007/s11336-024-09967-4>
10. Gökdemir, F., & Yılmaz, T. (2023). Likert tipi ölçekleri kullanma, modifiye etme, uyarlama ve geliştirme süreçleri. *Journal of Nursology*, 26(2), 148–160. <https://doi.org/10.5152/JANHS.2023.22260>
11. Gravetter, F. J., & Wallnau, L. B. (2021). *Essentials of statistics for the behavioral sciences* (10th ed.). Cengage Learning.
12. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis* (8th ed.). Cengage Learning.
13. Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2022). Partial least squares structural equation modeling (PLS-SEM) using R: A workbook. Springer. <https://doi.org/10.1007/978-3-031-06692-1>
14. Iannario, M., Iodice D'Enza, A., & Romano, R. (2022). A hybrid approach for the analysis of complex categorical data structures: Assessment of latent distance learning perception in higher education. *Computational Statistics*, 37, 2611–2632.



<https://doi.org/10.1007/s00180-022-01272-x>

15. Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1217–1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
16. Kankaraš, M., & Capecchi, S. (2025). Neither agree nor disagree: Use and misuse of the neutral response category in Likert-type scales. *METRON*, 83, 111–140. <https://doi.org/10.1007/s40300-024-00276-5>
17. Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.
18. Kramer, M., & Weldon, P. J. (2023). Constructing composite scores for contemporaneous behaviors: A comparison of four approaches. *Ethology*, 129, 489–497. <https://doi.org/10.1111/eth.13382>
19. Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of response options, reliability, validity, and potential bias in the use of the Likert scale education and social science research: A literature review. *International Journal of Educational Methodology*, 8(4), 625–637. <https://doi.org/10.12973/ijem.8.4.625>
20. Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1–55.
21. McClure, K., Ammerman, B. A., & Jacobucci, R. (2024). On the selection of item scores or composite scores for clinical prediction. *Multivariate Behavioral Research*, 59(3), 566–583. <https://doi.org/10.1080/00273171.2023.2292598>
22. McNeish, D. (2023). Dynamic fit index cutoffs for categorical factor analysis with Likert-type, ordinal, or binary responses. *Psychological Methods*, 28(6), 1061–1075. <https://doi.org/10.1037/met0001213>
23. Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
24. Olsson, U. (2022). Power properties of ordinal regression models for Likert type data. *Practical Assessment, Research, and Evaluation*, 27, Article 6. <https://doi.org/10.7275/er28-5992>
25. Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). McGraw-Hill Education.
26. Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson.
27. Sekaran, U., & Bougie, R. (2020). *Research methods for business: A skill-building approach* (8th ed.). Wiley.
28. Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
29. Tutz, G. (2021). Hierarchical models for the analysis of Likert scales in regression and item response analysis. *International Statistical Review*, 89(1), 18–35. <https://doi.org/10.1111/insr.12396>
30. Xu, S., & Liu, Y. (2022). Characterizing sampling variability for item response theory scale scores in a fixed-parameter calibrated projection design. *Applied Psychological Measurement*, 46(6), 509–528. <https://doi.org/10.1177/01466216221108136>