

# A streamlined approach to scale adaptation: Enhancing validity and feasibility in educational measurement

Yiran Du 

Institute of Cognitive Neuroscience, University College London, London, United Kingdom

Received: August 4, 2024 / Accepted: September 13, 2024 / Published Online: September 14, 2024  
© Pioneer Publications LTD 2024

## Abstract

In educational research, the adaptation of measurement scales to different contexts is necessary but often lacks thorough validation, leading to unreliable and invalid results that compromise research integrity and the effectiveness of educational policies. In this paper, I highlight these issues and advocate for elevated research standards and rigorous validation processes. I propose a practical validation approach that balances feasibility and rigor, including initial content validity checks, pilot testing with small samples, simplified statistical validation, and cross-validation when possible. While practical, this method has limitations, such as potential biases and incomplete psychometric assessments. Despite these, it enhances the reliability and validity of adapted scales, contributing to more robust and trustworthy educational research. By prioritizing meticulous validation, I believe researchers can uphold their ethical responsibilities and advance educational practices and policies grounded in robust evidence.

**Keywords** scale adaptation, validation process, educational measurement, reliability, psychometrics

## 1. Introduction

In educational research, measurement scales are indispensable tools for assessing a wide range of constructs, from student attitudes and perceptions to learning outcomes and teacher effectiveness (Finch & French, 2019). In this study, the term "scale" specifically refers to multi-item Likert scales, which are widely used in questionnaires to measure constructs such as attitudes, perceptions, and behaviors. These scales consist of multiple statements designed to assess a specific construct, with respondents indicating their level of agreement on a Likert-type scale, typically ranging from 1 (strongly disagree) to 5 (strongly agree). For instance, a multi-item Likert scale used to measure student motivation might include items such as "I am motivated to study regularly," "I find my coursework engaging," and "I set clear academic goals for myself." The scores from individual items are then combined to provide an overall measure of the construct. These scales are often adapted to ensure they remain relevant and applicable across diverse contexts, involving processes such as translating scales into different languages, modifying items to reflect cultural nuances, or adjusting the format to align with the target population. While these adaptations are necessary to enhance the scales' effectiveness, they introduce potential challenges, particularly concerning the scales' reliability and validity (Bateson & Martin, 2021).

Despite the critical role of these adaptations, a troubling trend has emerged: many researchers neglect

thorough validation processes. This oversight can lead to the use of scales that produce unreliable and invalid results, compromising the integrity of research findings. The implications of using poorly validated scales are far-reaching, affecting not only the quality of academic research but also the practical applications of this research in educational policy and practice. Unvalidated scales can result in erroneous conclusions, misguiding subsequent research and leading to ineffective or detrimental educational interventions (Hopkins, 1998).

Given the importance of scale validation, I believe it is imperative for educational researchers to elevate their research standards and avoid arbitrary adaptations of scales. We must adopt rigorous and systematic approaches to ensure the reliability and validity of adapted scales. Detailed validation processes, however, are often time-consuming and resource-intensive, posing a significant challenge for researchers. Therefore, in this article, I propose a practical and efficient approach to scale validation that balances thoroughness with feasibility. By following a streamlined validation process, researchers can maintain the quality of their measurement tools without the need for extensive separate studies.

The subsequent sections of this article will outline the current state of scale adaptation, the consequences of neglecting validation, and a step-by-step guide for efficiently validating adapted scales. Additionally, the limitations of this streamlined approach will be discussed to provide a comprehensive understanding of its applicability and potential shortcomings. Through this approach, I aim to help educational researchers enhance

the credibility and impact of their findings, ultimately contributing to more robust and reliable educational research. By committing to rigorous validation, we can uphold our ethical responsibilities and advance educational practices and policies grounded in trustworthy evidence.

## 2. The Current State of Scale Adaptation

Scale adaptation is a widespread practice in educational research, driven by the necessity to tailor measurement tools to specific contexts, populations, and research objectives (Stewart et al., 2012). Researchers frequently modify existing scales to enhance their relevance and accuracy. This adaptation process often involves translating items into different languages, adjusting them to reflect cultural nuances, or reformatting the scale to suit the characteristics of the target population. These modifications aim to ensure that the scale accurately measures the intended constructs within the new context, thereby improving the overall validity of the research.

Despite the widespread use of adapted scales, there is a concerning trend of researchers neglecting thorough validation processes. One of the most significant oversights is the failure to account for cultural differences adequately. When scales are adapted without considering the cultural context in which they will be used, the items may be interpreted differently by different populations, leading to skewed and unreliable results (Rossier & Duarte, 2019). This oversight is particularly problematic in cross-cultural research, where the goal is often to compare findings across different cultural groups.

Another common oversight in the adaptation of scales is the lack of pilot testing. Pilot testing is a crucial step that helps identify and rectify issues related to item clarity and relevance before the main study is conducted (Morgado et al., 2018). Without pilot testing, researchers risk using scales that do not function as intended, which can compromise the validity of their findings. This step ensures that any potential problems with the adapted scale are addressed early on, thereby enhancing the overall quality of the research.

Additionally, many researchers fail to perform necessary statistical validations to confirm the psychometric properties of their adapted scales. Essential validation steps, such as factor analysis, reliability analysis, and validity testing, are often overlooked. This neglect can result in the use of scales that are not reliable or valid, leading to erroneous conclusions. The primary consequence of this oversight is the production of unreliable and invalid research findings, which can significantly undermine the integrity of the research.

Moreover, many researchers do not provide the adapted versions of their scales or clearly articulate the modifications they have made. This lack of transparency can hinder the replicability and scrutiny of research findings, as other researchers cannot fully understand or critique the adaptation process. Sharing detailed information about scale adaptations, including the rationale and specific changes made, is essential for

advancing the field and ensuring the robustness of research outcomes.

## 3. Consequences of Neglecting Validation

The failure to rigorously validate adapted scales has significant and far-reaching consequences. The most immediate impact is the production of unreliable and invalid research findings (Clark-Carter, 2019). When scales are not properly validated, the data collected may not accurately reflect the constructs being measured. This misalignment leads to erroneous conclusions, which can misguide subsequent research and misinform educational practices and policies. The credibility of the research is compromised, and the intended contributions to the academic field are diminished.

One of the primary consequences of using unvalidated scales is the risk of generalizing incorrect findings (Ramsey & Schafer, 2013). Researchers rely on the data collected from these scales to draw conclusions and make recommendations. If the scales are not reliable or valid, the conclusions drawn from the data are fundamentally flawed. This can lead to a cascade of misinformation, as other researchers may build upon these flawed findings, perpetuating errors and potentially leading to widespread misconceptions within the field.

In the context of educational policy and practice, the implications of using invalid scales are particularly concerning. Research findings often inform policy decisions and educational interventions. When these findings are based on data from unvalidated scales, the resulting policies and interventions may be ineffective or even detrimental (Lingard, 2013). For example, educational programs designed to address student needs based on flawed data may fail to achieve their objectives, wasting resources and potentially causing harm to students and educators.

Moreover, the ethical responsibility of researchers is called into question when validation is neglected (Frank et al., 2024). Researchers have an obligation to ensure that their measurement tools are accurate and reliable, as the integrity of their findings directly impacts the people and systems they study. Neglecting validation not only undermines the quality of the research but also violates this ethical duty, potentially leading to mistrust in the research community and among the public.

## 4. A Practical and Efficient Approach for Validation

Validating an adapted scale in educational research can be streamlined to balance thoroughness with feasibility. This approach ensures that adapted scales maintain their reliability and validity without necessitating extensive, time-consuming studies. Here is a step-by-step guide to efficiently validate an adapted scale (see Table 1).

Begin with an expert review and focus groups to establish initial content validity. Assemble a small group of experts (3-5 individuals) familiar with the construct and

cultural context to review the adapted items for relevance, clarity, and cultural appropriateness. Collect their feedback, noting any suggested revisions or concerns. Additionally, conduct focus group discussions with a small, representative sample of the target population (6-10 participants). Engage participants in discussions about each item, focusing on clarity, relevance, and cultural understanding. This step provides valuable insights and helps identify potential issues early in the process.

Next, conduct a pilot study with a small, representative sample of the target population (20-30 participants). Administer the adapted scale and collect responses. Analyze descriptive statistics to identify items with unusual response patterns or high non-response rates. Perform a preliminary reliability analysis by calculating Cronbach's alpha for the entire scale and any subscales. Identify items with low item-total correlations or that significantly reduce the overall reliability (Cronbach's alpha < 0.70). This pilot testing phase is crucial for detecting and addressing initial problems with the adapted scale.

Following the pilot study, conduct an exploratory factor analysis (EFA) using software like SPSS or R. Examine the factor loadings to ensure that items load onto the expected factors, and revise or remove items with low factor loadings (typically < 0.40) or those that do not fit well with other items. This step helps verify the underlying structure of the scale and ensures that it measures the intended constructs. After refining the scale based on the EFA results, perform a reliability analysis again. Calculate Cronbach's alpha for the revised scale and any subscales to reassess internal consistency. Aim for a minimum threshold of Cronbach's alpha > 0.70 to indicate acceptable reliability (Harrison et al., 2022). This second round of reliability testing confirms that the refined scale maintains or improves its internal consistency. Based on the feedback from the expert review, focus groups, pilot testing, and EFA, make necessary revisions to the scale items. Document each change and the rationale behind it

to maintain transparency.

If resources allow, conduct a cross-validation study with a new, small sample from the target population (another 20-30 participants). Administer the revised scale and collect responses, comparing the results with the initial pilot study to check for consistency in reliability and factor structure. Combining quantitative and qualitative methods can provide a comprehensive understanding of the scale's performance. Use qualitative feedback from the focus groups to contextualize the quantitative findings and reassess any unexpected results with qualitative insights to understand underlying issues.

Make final adjustments based on the cross-validation and mixed-methods feedback, ensuring the scale is clear, culturally appropriate, reliable, and valid. Provide detailed documentation of the entire adaptation and validation process, including changes made and reasons for each change. Share the adapted scale and validation results with the research community to enhance transparency and replicability. By following these steps, educational researchers can efficiently validate their adapted scales, ensuring both practicality and rigor in their validation processes.

While this streamlined approach to validating adapted scales offers a practical balance between thoroughness and feasibility, it has several limitations. The reliance on small sample sizes in pilot testing and cross-validation may not capture the full diversity of the target population, potentially overlooking subtle but important variations. Additionally, expert reviews and focus groups, while insightful, may be subject to bias and may not fully represent the broader population's perspectives. The simplified statistical validation techniques, although useful, may not be as comprehensive as full-scale validation studies, potentially leaving some psychometric issues unaddressed. Lastly, the iterative refinement process, though efficient, might still miss deeper, systemic issues that more extensive validation efforts could uncover.

**Table 1.** Key Steps in the Streamlined Scale Adaptation and Validation Process

Step	Description	Purpose
Step 1: Initial Content Validity Check	Conduct expert reviews and focus groups with representatives from the target population to assess item clarity, relevance, and cultural appropriateness.	Ensures that items are relevant and understandable within the new context.
Step 2: Pilot Testing with Small Samples	Administer the adapted scale to a small, representative sample (20-30 participants). Analyze response patterns and reliability (e.g., Cronbach's alpha).	Identifies problematic items and assesses initial reliability of the scale.
Step 3: Simplified Statistical Validation	Conduct exploratory factor analysis (EFA) to verify the factor structure and internal consistency of the scale. Refine items as needed.	Confirms that the scale measures the intended constructs and refines items based on statistical analysis.
Step 4: Cross-Validation (Optional)	If possible, test the revised scale with a different sample from the target population (another 20-30 participants).	Provides additional evidence of the scale's reliability and consistency.
Step 5: Finalization and Reporting	Make final adjustments based on pilot and cross-validation results. Document the process and share the adapted scale and validation results.	Ensures transparency, allowing other researchers to replicate or critique the adaptation process.

## 5. Discussion

The present study underscores the necessity of rigorous validation processes in the adaptation of educational measurement scales. Despite the widespread practice of scale adaptation to suit diverse research contexts, there is a concerning trend of neglecting thorough validation. This oversight has significant implications, including the production of unreliable and invalid research findings, which can mislead subsequent research and result in ineffective or even harmful educational policies and interventions.

One of the primary issues identified is the inadequate consideration of cultural differences when adapting scales. Without proper validation, adapted scales may fail to account for the nuances of different cultural contexts, leading to misinterpretations and skewed results. This problem is particularly acute in cross-cultural research, where the aim is often to compare findings across diverse groups. The lack of pilot testing is another critical oversight. Pilot tests are essential for identifying and rectifying issues related to item clarity and relevance, yet many researchers omit this step, risking the use of scales that do not function as intended. Additionally, the absence of necessary statistical validations, such as factor analysis, reliability analysis, and validity testing, further compromises the integrity of adapted scales. Without these validations, researchers cannot confirm the psychometric properties of their scales, leading to erroneous conclusions. The lack of transparency in reporting adapted versions of scales and the modifications made is another major concern. This opacity hinders the replicability and scrutiny of research findings, as other researchers cannot fully understand or critique the adaptation process.

To mitigate these issues, this paper proposes a streamlined yet rigorous approach to scale validation. This approach includes initial content validity checks through expert reviews and focus groups, pilot testing with small samples, simplified statistical validation, and cross-validation if resources allow. By implementing these steps, researchers can ensure that their adapted scales are both reliable and valid, thereby enhancing the overall quality and credibility of their research findings.

However, this streamlined approach is not without its limitations. The reliance on small sample sizes in pilot testing and cross-validation may not capture the full diversity of the target population, potentially overlooking subtle but important variations. Expert reviews and focus groups, while insightful, may be subject to bias and may not fully represent the broader population's perspectives. The simplified statistical validation techniques, though useful, may not be as comprehensive as full-scale validation studies, potentially leaving some psychometric issues unaddressed. Additionally, the iterative refinement process, although efficient, might still miss deeper systemic issues that more extensive validation efforts could uncover. To mitigate these limitations, researchers can adopt several targeted strategies. First, expanding the pilot and cross-validation sample sizes to include at least 50-100 participants, or more, will improve the

representativeness of the target population and help capture subtle variations. Additionally, researchers should actively recruit participants from diverse demographic backgrounds, including different age groups, educational levels, and cultural contexts, to ensure the scale's generalizability. For expert reviews and focus groups, selecting experts with diverse areas of expertise and cultural backgrounds, and using blinded reviews where experts are unaware of the study's specific aims, can reduce bias and provide a more well-rounded assessment of the scale. Furthermore, while exploratory factor analysis (EFA) offers an initial understanding of the scale's structure, following up with confirmatory factor analysis (CFA) can provide a more rigorous test of the hypothesized factor structure, addressing potential psychometric issues. Lastly, incorporating multiple iterations of testing with progressively larger and more varied samples can help identify and address deeper, systemic issues that may not surface in early stages, ensuring a more thorough and reliable validation process.

Despite these limitations, the proposed approach offers a practical and balanced solution for researchers, ensuring the integrity of adapted scales without requiring extensive, resource-intensive studies. By adopting these methods, researchers can produce more reliable and valid research findings, contributing to the advancement of educational research and the development of effective educational practices and policies. Ultimately, ensuring the validity of adapted scales is not only a methodological necessity but also an ethical responsibility that researchers must uphold to maintain the trust and efficacy of educational research.

To conclude, it is imperative for educational researchers to elevate their research standards and avoid arbitrary adaptations of scales. Thorough validation is not merely a procedural step but a foundational element that ensures the reliability and validity of research findings. Neglecting this critical process can lead to the dissemination of inaccurate and misleading results, ultimately compromising the integrity of the research and its practical applications. By committing to rigorous validation, researchers uphold their ethical responsibilities and contribute to the advancement of educational practices and policies that are grounded in robust and trustworthy evidence. Let us advocate for a research culture that prioritizes meticulous validation, fostering greater confidence and credibility in the scientific community.

**Yiran Du** is an MRes student in Cognitive Neuroscience at University College London. He holds an MEd degree in Human Development and Education from Harvard University. He is interested in verbal irony comprehension, computer-assisted language learning, and autism.  
Email: [yiran.du.24@ucl.ac.uk](mailto:yiran.du.24@ucl.ac.uk)

### References

- Bateson, M., & Martin, P. R. (2021). *Measuring behaviour: An introductory guide* (4th ed). Cambridge University Press.
- Clark-Carter, D. (2019). *Quantitative psychological*

- research: *The complete student's companion* (4th Edition). Routledge.
- Finch, W. H., & French, B. F. (2019). *Educational and psychological measurement*. Routledge.
- Frank, M., Braginsky, M., Cachia, J., Coles, N., Hardwicke, T., Hawkins, R., Mathur, M., & Williams, R. (2024). *Experimentology: An open science approach to experimental psychology methods*. MIT Press.  
<https://doi.org/10.7551/mitpress/14810.001.0001>
- Harrison, V., Kemp, R., Brace, N., & Snelgar, R. (2022). *SPSS for psychologists* (Seventh edition). Bloomsbury Academic.
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed). Allyn & Bacon.
- Lingard, B. (2013). The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education*, 54(2), 113–131.  
<https://doi.org/10.1080/17508487.2013.781515>
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2018). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(1), 3.  
<https://doi.org/10.1186/s41155-016-0057-1>
- Ramsey, F. L., & Schafer, D. W. (2013). *The statistical sleuth: A course in methods of data analysis* (3rd ed). Brooks/Cole, Cengage Learning.
- Rossier, J., & Duarte, M. E. (2019). Testing and assessment in an international context: Cross- and multi-cultural issues. In J. A. Athanasou & H. N. Perera (Eds.), *International handbook of career guidance* (pp. 613–637). Springer International Publishing. [https://doi.org/10.1007/978-3-030-25153-6\\_28](https://doi.org/10.1007/978-3-030-25153-6_28)
- Stewart, A. L., Thrasher, A. D., Goldberg, J., & Shea, J. A. (2012). A framework for understanding modifications to measures for diverse populations. *Journal of Aging and Health*, 24(6), 992–1017.  
<https://doi.org/10.1177/0898264312440321>

### Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2024 Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

## Call for Papers

### Submit via <https://jlt.ac/>

#### Areas of Interest:

Language teaching intervention and experiments; Curriculum development; Language teacher education; Bilingual education; New technologies in language teaching; Testing, assessment, and evaluation; Educational psychology, and more.

#### We accept the following types of submission:

1. Research article: (6,000 to 8,000 words)
2. Review: (3,000 to 8,000 words)
3. Book review: (up to 3,000 words)
4. Features: (3,000 to 8,000 words)

Scan to submit your articles\* & read more articles for free.

\*Article Processing Charges Apply.



Contact: [editor@jlt.ac](mailto:editor@jlt.ac)