Puissant (ISSN print: 2719-0153; ISSN online: 2719-0161)
Vol. 7, pp. 3083-3092
https://puissant.stepacademic.net

**Position Paper**

# The Categorical Ceiling Tax: Interpreting Intervention Outcomes in Educational Research

Mhel Cedric D. Bendo
College of Business Administration, Polytechnic University of the Philippines, Cavite, Philippines
cedricbends@gmail.com
ORCID: 0009-0004-6873-3910

*Date received*: January 26, 2026
*Date received in revised form*: February 4, 2026; February 8, 2026
*Date accepted*: February 14, 2026

Recommended citation:

Bendo, M. C. D. (2026). The categorical ceiling tax: Interpreting intervention outcomes in educational research. *Puissant*, *7,* 3083-3092.

**Abstract**

This position paper formalizes the concept of the Categorical Ceiling Tax (CCT), an interpretive limitation observed in intervention studies that employ quantitative scoring rubrics to classify proficiency levels. The CCT describes a form of structural resistance that occurs when a population mean approaches, but does not cross, a predefined categorical threshold, even when measurable quantitative gains are documented. Using pre–post score data from a one-group design involving the total population of N = 494 Grade 7 students, this paper illustrates how categorical reporting obscured a documented positive mean shift ($\Delta\bar{x}$ = +0.20), while the population classification remained at the "Instructional" level. This paper argues that categorical transitions, when interpreted independently, may misrepresent population-level progress in large samples. It further proposes that changes in the deficiency floor—defined as the proportion of learners classified at the Frustration level—be reported as a complementary interpretive indicator alongside categorical outcomes. These observations provide applied insights into the interpretation of outcomes generated through technology-supported assessment and reporting systems commonly used in contemporary educational settings.

*Keywords* – Categorical Ceiling Tax, Educational Measurement and Evaluation, Categorical Reporting and Cut Scores, Reading Intervention Interpretation, Deficiency-Floor Indicators

## INTRODUCTION

It is common for intervention studies within the education literature to apply rating scales, where raw intervention data (continuous) are measured and ranked according to predetermined standards that separate the data into distinct, defined categories (often including Frustration, Instructional, and Independent). Such data are critical for summarizing intervention results, but they also pose interpretive challenges when used to describe outcomes in large-scale interventions. The issue, termed the Categorical Ceiling Tax (CCT), develops when the average baseline score of a population lies just below the threshold defining the next level of proficiency. Even when an intervention is associated with a statistically significant positive shift in average score, it is possible that the population mean will not surpass the categorical cut point (Zieky & Perie, 2006). In such cases, the categorical descriptor remains unchanged, despite documented quantitative movement.

The phenomenon described here aligns with several well-documented measurement concerns in educational research, including ceiling effects, loss of information due to categorization, and sensitivity to cut scores (Stevens, 1946; Cohen et al., 2003; Taber, 2018; MacCallum et al., 2002; Royston et al., 2006; Lord, 1980). Ceiling effects typically refer to score saturation at the upper bound of an instrument, whereas categorization loss refers more broadly to the reduction of informational precision when continuous data are collapsed into discrete levels. The Categorical Ceiling Tax, however, differs in emphasis and application. Rather than focusing on psychometric limitation or score compression, the CCT refers specifically to an interpretive penalty at the population level, wherein substantial quantitative gains do not translate into categorical advancement due to proximity to predefined cut points. This penalty becomes especially salient in large-scale or census-based studies, where population means exhibit strong inertia against categorical transition despite widespread individual change.

This position paper does not introduce a new psychometric principle; rather, it formalizes an applied interpretive limitation frequently encountered in educational reporting but often under-articulated in evaluation practices. By naming and illustrating this limitation, the paper aims to support clearer and more ethically grounded interpretation of intervention outcomes in applied educational research. To illustrate this methodological concern, the paper draws on empirical data collected from a one-group study conducted over a one-month period among a total population of N = 494 Grade 7 students (Bendo et al., 2026). Prior research documents changes in reading performance and affective learning challenges in the Philippine educational context (Hutamares, 2023; Bernardo, 2010). However, in the present illustration, the population descriptor remained at the "Instructional" level ($\bar{x}$ = 2.19) despite documented score movement, providing a clear example of the CCT.

This position paper offers an applied methodological contribution regarding the interpretation of intervention outcomes, emphasizing deficiency-floor reduction rather than categorical ceiling transition as a primary interpretive indicator. This reframing supports more structurally coherent reporting practices, particularly in contexts where universal mastery within a short intervention period is unrealistic (Taber, 2018). The phenomenon under observation reflects a limitation associated with large-sample analyses. When census approaches are used, population means may be resistant to categorical movement across broad thresholds, even when individual-level change is present. This issue reflects a structural feature of categorical classification rather than a property of the intervention itself. In applied educational settings, where reaching universal mastery ($\bar{x} > 2.33$) within limited timeframes is often infeasible, such interpretive constraints warrant careful consideration.

Traditional reliance on categorical transition as the sole indicator of intervention outcome does not account for the practical importance of reducing severe learning deficiencies. In resource-constrained educational contexts, a reduction in the proportion of learners at the Frustration level represents a substantively meaningful population shift. Accordingly, this paper formalizes the CCT to support reporting practices that distinguish documented quantitative change from categorical classification constraints.

## METHODOLOGY

### *Techniques and Data Source*

The methodological analysis draws on data from a one-group pretest–posttest design study (Bendo et al., 2026) among the targeted participants, that is, the total number of participants, N = 494, comparable to the rules for quantitative design study requirements (Calmorin & Calmorin, 2007; Shields & Rangarajan, 2013).

This position paper utilizes a standardized reading comprehension test appropriate for the Philippine Informal Reading Inventory (Phil-IRI) design. The design of the study must guarantee the quality of the instrument and the precision of the data measurement (Taber, 2018). This conformance to a census of the complete Grade 7 population (N = 494) supports consistency in the observed score patterns and reduces sampling-related variability when examining how quantitative results are represented through categorical proficiency scales. Moreover, standardization and validation of instruments such as those within the Phil-IRI framework are key to establishing instrument score reliability before a formal critique of the interpretive rubric is undertaken.

The pretest–posttest structure provides a documented score contrast that is used illustratively to examine how categorical reporting frameworks interact with observed quantitative change. Such documentation is necessary as part of a methodological argument that seeks to examine how quantitative designs interact with categorical classification systems.

The purpose of the present analysis is not to re-evaluate the effectiveness of the intervention itself, but to examine how commonly used categorical reporting practices may constrain the interpretation of documented quantitative gains in applied educational contexts. This study adopts an applied mixed-methods orientation, using quantitative intervention data as illustrative evidence to examine a methodological issue in educational reporting.

### *Analytic Rationale*

The analytic focus of this position paper is deliberately descriptive and interpretive rather than inferential. Mean score changes and categorical descriptors are examined to illustrate the paper's position that substantial quantitative gains may remain invisible under categorical reporting frameworks. This approach aligns with the position paper's purpose of examining interpretive limitations in applied educational research rather than estimating causal effects or population parameters.

Mean scores were summarized using the weighted mean formula and interpreted using the established proficiency scale.

### *Metrics and Interpretation*

Mean scores were summarized using the weighted mean formula and interpreted using the established proficiency scale:

Table 1. Scoring Rubric and Proficiency Scale (3-Point Likert Scale)

| Scale Point | Proficiency Level | Score Range |
|---|---|---|
| 3 | Independent | 2.34 – 3.00 |
| 2 | Instructional | 1.67 – 2.33 |
| 1 | Frustration | 1.00 – 1.66 |

### *Focus and Analysis*

The central theme of this study will revolve around the comparison of the scale of the quantitative gain, $\Delta\bar{x}$, and the categorical statement that aptly defines the total population. This will also cover the accurate depiction of the existing situation (Calmorin, 2007).

## RESULTS

This analysis highlights a divergence between the progress made quantitatively and the ultimate categorization given:

Table 2. Pretest–Posttest Outcomes and Categorical Classification (N = 494)

| Metric | Pre-Test Score | Post-Test Score | Change ($\Delta\bar{x}$) | Resulting Category |
|---|---|---|---|---|
| Overall Weighted Mean | 1.99 | 2.19 | +0.20 | Instructional |
| Deficiency Floor (Frustration %) | 22.27% | ≈ 6.48% | Marked Reduction | Mitigation Achieved |

- **The Tax in Effect:** The population weighted mean increased from 1.99 (pre-test) to 2.19 (post-test), representing a measurable gain ($\Delta\bar{x}$ = +0.20). However, because the post-test mean did not exceed the categorical cut point of 2.33 required for the "Independent" level, the overall population classification remained at "Instructional." This pattern illustrates the Categorical Ceiling Tax, wherein categorical thresholds introduce a zone of interpretive insensitivity such that quantitative gains are not reflected in categorical outcomes. In large populations (N = 494), the magnitude of change required to cross a categorical boundary may exceed the realistic scope of short-term interventions. As a result, categorical stability may coexist with documented linear improvement.

- **The Untaxed Indicator:** By contrast, the Deficiency Floor showed a marked change. The total group of participants (N = 494) reduced the level of Frustration from 22.27% (pre-test) to approximately 6.48% (post-test). Additionally, the sections of both Punongbayan and Almeda recorded 0% Frustration, while one section of Punongbayan exhibited a categorical shift from Instructional to Independent ($\bar{x}$ = 2.33–2.51). This distributional change illustrates a population movement that is not constrained by categorical ceiling thresholds. As such, Frustration-level reduction provides additional context for interpreting categorical outcomes in large-sample analyses and highlights the importance of reporting deficiency-floor indicators alongside categorical classifications (Embretson & Reise, 2000).

## DISCUSSION

This position paper demonstrates that reliance on categorical proficiency labels in large-scale intervention studies can limit the visibility of documented quantitative change. In the present illustration, the Instructional category label persisted despite a population-level increase in mean score ($\Delta\bar{x}$ = 0.20). The Categorical Ceiling Tax represents an applied interpretive concept that highlights a structural limitation in the interpretation of empirical results when standardized categorical language is used. In this case, categorical classification did not reflect the documented magnitude of quantitative movement observed across the population.

For instance, when the average of a large sample begins close to a categorical level (as 1.99 is close to 2.33), the force of inertia needed for a categorical outcome may well exceed the boundaries of a brief intervention session. This design weakness could potentially result in the misjudgment of the potency of their intervention efforts, thus undermining the motivation to document successful initiatives that only brought about large, albeit not categorical, improvements (Hutamares, 2023).

This drawback brings about the need for a new approach when it comes to the dissemination of intervention results:

- **Shift Focus from the Ceiling:** Authors need to resist the temptation of categorical shift becoming the success criteria when writing large sample intervention studies

- **Feature the Deficiency Floor**: The strongest, most ethical, and potent measure of success, especially for resource-deprived schools, is the measurable lessening of the deficiency floor, that is, the percentage of reduction of the number of students at the Frustration level. The removal or substantial reduction of the need for emergency assistance is the clearest, nontaxed measure of success, which is free from the inherent bias of the proficiency scale's higher ceilings.

The CCT notion is an essential measurement criticism, which confirms that the design of methodology should express practical and ethical objectives of intervention (Taber, 2018; American Educational Research Association [AERA], 2011; American Psychological Association [APA], 2020). The result also confirms comparable evidence regarding the effectiveness of reading intervention programs (Hutamares, 2023), but also redefines how the effectiveness of such interventions is to be quantified and communicated to their stakeholders. Formalizing the CCT, however, gives the researchers a new model of making certain that the hard-to-reach academic settings do not imply that the progress has been made.

***Practical Reporting Implications for Educational Research***

To mitigate the interpretive limitations associated with categorical reporting in large-scale interventions, the following applied reporting practices are recommended for educational researchers and practitioners:

1. Continuous gains (e.g., mean score change) should always be reported alongside categorical descriptors.

2. Deficiency-floor indicators, such as the percentage of learners at the Frustration level, should be explicitly reported and interpreted as primary outcomes where relevant.

3. Cases in which statistically or practically meaningful gains occur without categorical transition should be explicitly flagged rather than interpreted as null effects.

4. Evaluation conclusions should avoid binary judgments of "effective" or "ineffective" when outcomes fall near categorical cut points.

These practices require no additional data collection and can be readily adopted within existing school-based evaluation systems.

***Practical Insights for Researchers***

This paper provides critical insights for educational practitioners. Specifically, it suggests that relying solely on categorical shifts (e.g., Frustration to Instructional) may lead to the premature termination of successful programs. Researchers are encouraged to adopt 'untaxed' indicators, such as raw score growth within categories, to ensure a more equitable representation of student progress.

## CONCLUSIONS AND RECOMMENDATIONS

Finally, the Categorical Ceiling Tax shows that there is a major methodological drawback in reporting the intervention outcomes with the help of proficiency rubrics. Educational survey researchers are advised to build on this framework so as to give a more insightful, precise, and scientifically justifiable character of the treatment effects by making quantifiable reduction of deficiency the most pertinent measure of successful implementation of a program (Bendo et al., 2026). This methodological change is justified by the high ethical consequences of the survey reporting. The misinterpretation of a $\Delta\bar{x}$ gain as a structural failure as a result of CCT may have a direct impact on funding and resources allocation toward specific education in the future. The CCT becomes a systematic hindrance to the apparent success of the efforts, although of the highest effect, which, though very dynamic, are limited in number.

This position paper thus supports a system of reporting standard more transparent through the inclusion of both granular statistical data and categorical designation, thus offering both dual-metric evaluation of program effectiveness. This two-metric method, in which the combined weighted mean variation should be observed, and the cutback percentage of the deficiency floor, is a scientifically fair methodology of assessing the progress in heterogeneous populations. This position papers plainly acknowledges that in areas like the remediation of literacy, a shift of a student out of a failing group is much more significant and unmasked measure of achievement than an overall increase in group mean.

The proposed framework offers a more informed and ultimately more justifiable view on the intervention effects, and the arbitrary structural punishments of the conventional

rubrics. Future research ought to concentrate in putting the CCT model to test in the various domains of standardized assessment so as to ensure its generalizability across the various situations of quantitative survey.

## ACKNOWLEDGEMENT

The author acknowledges the administration and Grade 7 teachers of Public National Integrated School for their cooperation during the original data collection phase of the reading intervention study from which the illustrative data used in this paper were derived. Appreciation is also extended to the students who participated in the original study.

## FUNDING

This study did not receive funding from any public, commercial, or non-profit institution.

## DECLARATIONS

### *Conflict of Interest*

The author declares no conflict of interest.

### *Informed Consent*

Informed consent was obtained from all participants as part of the original reading intervention study conducted at a public national integrated school. The present paper uses anonymized and aggregated data solely for methodological and interpretive analysis.

### *Ethics Approval*

Ethical approval for data collection was secured during the original intervention study in accordance with institutional and school-based research protocols. The current manuscript does not involve new data collection and relies exclusively on secondary analysis of previously approved data.

## REFERENCES

American Educational Research Association. (2011). Code of ethics. *Educational Researcher, 40*(3), 145–156. https://doi.org/10.3102/0013189X11410403

American Psychological Association. (2020). *Publication manual of the American Psychological Association (7th ed.)*. American Psychological Association.

Bendo, M. C. D., Angue, D. D., Buhay, J., Dimaisip, J. A., Lobos, T. L., Pahimnayan, E. G., Piedad, R. B., France Lawrence Pulgo, F. L., Ramos, K., & Yabut, W. A. (2026). Enhancing Grade 7 reading proficiency through an English reading program intervention in high school. *Journal of Research, Innovation, and Strategies for Education (RISE), 3*(2), 90–110. https://doi.org/10.70148/v3i3.7

Bernardo, A. B. I. (2010). Extending self-regulated learning theory: Filipino students' academic emotions and achievement goals. *The Asia-Pacific Education Researcher, 19*(2), 259–271. https://doi.org/10.3860/taper.v19i2.1621

Calmorin, L. P., & Calmorin, M. A. (2007). *Research methods and thesis writing (2nd ed.)*. Rex Book Store.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203774441

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Hutamares, R. (2023). The effectiveness of Brigada Pagbasa in improving the reading comprehension of junior high school students at Gumaca National High School. *Psychology and Education: A Multidisciplinary Journal, 9.* https://doi.org/10.5281/zenodo.7973782

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum Associates.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine, 25*(1), 127–141. https://doi.org/10.1002/sim.2331

Shields, P. M., & Rangarajan, N. (2013). A playbook for research methods: Integrating conceptual frameworks and project management. New Forums Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680. https://doi.org/10.1126/science.103.2684

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-018-9765-4

Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Educational Testing Service.

## Author's Biography

Mhel Cedric D. Bendo is a student researcher specializing in educational psychology, curriculum and instruction, educational measurement and evaluation, psychometrics, and educational technology. He is pursuing a Bachelor of Science in

Entrepreneurship at the Polytechnic University of the Philippines. As of 2026, he holds an honorary appointment as an International Editorial Board Member for SIAR Publications, contributing to peer review, editorial guidance, and interdisciplinary journal development. He also serves as an invited peer reviewer for international peer-reviewed journals in online and distance learning, educational technology, educational evaluation, information literacy, interdisciplinary undergraduate research, and health and science education, including journals indexed in Web of Science (ESCI), Scopus (Q1–Q3, SCImago), ERIC, and DOAJ, supporting rigorous and ethically grounded interdisciplinary scholarship.