

Date

Ref: xxx

Editor's name  
Name of Journal  
Address

Dear Editor,

Thank you for inviting us to revise and to resubmit our manuscript entitled, "xxx" (Provide Manuscript Number) for publication in xxx. Attached is our resubmission.

We have attempted to address all of the comments. Attached are descriptors of what we have changed. We are hopeful that we have corrected and clarified to the extent that the manuscript is ready for publication. Because of the suggestions made, we believe that our manuscript has been strengthened.

Sincerely,

Name of Corresponding Author, Qualifications  
Title  
Affiliation  
Email Address

Enclosure

## **A Summary of Changes to the Manuscript**

“Title of Manuscript” (Provide Manuscript Number)

Many thanks to the reviewers and editor for their excellent suggestions. We appreciate the diligent efforts of each reviewer and the editor and believe that their comments have helped to make our original manuscript much stronger. Although a few of the comments seemed to suggest simply further reflection, most either stated or implied specific revisions, and in nearly all of these cases we have complied. In the few instances in which we did not comply with the recommendations, we have provided a rationale for deciding against making the suggested change. We have listed below the ways in which we addressed the suggestions.

### **Editor**

We appreciate the editor’s thoughtful suggestions.

- (1) The editor requested, “Please ask the author to move the section on confidence intervals for the economic tests to an appendix, suitably signalled in the main text.” This is an excellent suggestion. We have created an Appendix and have moved the two sections on confidence intervals.
- (2) The editor requested the following: “Please ask the author to define and illustrate clinical significance and practical significance, including drawing a clear distinction between them, where pertinent with reference to literature.”

We thank the editor for asking us to clarify this relationship. As a result, we modified the text, inserting a section that includes definitions and limitations of practical, clinical, and statistical significance. We included statistical significance in this section to keep it balanced.

- (3) The editor asked us to that “make very much more explicit how economic significance is different from clinical and practical significance, so that, for example, it could not be seen as a sub-sub-set of clinical significance.” We agree that this is important to make this differentiation, and hope that the definitions and limitations sections will clarify this. Furthermore, we have added the paragraph below to the conclusion section.

Furthermore, economic significance differs from the three existing types of significance in three major aspects. First, economic significance is deterministic in its results; the results are based on behaviors or actions. Conversely, statistical significance is based on probabilities. Second, economic significance is reported in understandable language of monetary sums, whereas practical significance typically is represented in standard deviation units, which can be difficult for many consumers to understand. Finally, economic significance reflects the objectives of the researcher(s) and/or the stakeholder(s), whereas clinical significance is based on the subjective experience(s) of the person(s) included in the study.

## Reviewer 1

We are glad that the reviewer believes that our manuscript “makes a very useful contribution to the research and evaluation field” and “is very timely, clear, to the point, well argued and of a high quality, and the argument about the need for economic significance, coupled with how to calculate it, is well placed and the case is made persuasively.”

- (1) The reviewer states that “The paper is comparatively silent on the operationalization of important matters such as utility (p. 15), and it is not enough to mention ‘the value that the people involved place on the program’, as this skirts the issue”. With all due respect to the reviewer, without using real data, we are unable to operationalize further the concept of utility. As we noted in our paper,

The ranges of utility will change depending on the situation and the value that the people involved place on the program or intervention. If the people involved believe that the utility is very important, but not extremely important, they might change the ranges of the CU ESI. (p. 20)

Thus, as much as we would like explicitly to provide cut points for the CU ESI, we are unable to do so.

- (2) The reviewer states “*estimating* the amount of utility as the basis for subsequent analysis (p. 15) is weak and this needs to be addressed to make it firmer and more rigorous; even though the paper acknowledges this problem (p. 19) the solution is not adequately addressed. Similarly operationalizing the *worth* of an intervention (p. 18) needs to be unpacked more concretely, as the same issue applies here as for the utility issue”. With respect to estimating the utility, previously we stated that “Weaknesses of the CU ESI include that it can be very difficult to estimate consistent and accurate measures of usefulness, especially across people and different populations.” In an attempt to address this problem, as recommended by the reviewer, we inserted the following: “Therefore, it is imperative that researchers assess the score reliability (e.g., internal consistency, test-retest reliability) and validity (i.e., content-related validity, criterion-related validity, construct-related validity) of all measures of utility” (p. 22).
- (3) The reviewer states “the definition of economic significance as ‘economic value . . . ‘ (p. 8) is imprecise, i.e. the reader needs to have made very clear what constitutes ‘economic value’, as, otherwise, the whole definition, as it stands, is a tautology.” It is our belief that economic value changes from situation to situation. What constitutes value in one situation might be very different in another. We assume economic significance will be applied in many areas, thus, we cannot explicitly define “economic value.”
- (4) We agree with the reviewer that “it would be helpful to have a worked example of the cost-utility analysis (p. 15) and the nine steps on page 20 (even though the steps are explained subsequently, a worked example would clarify the matters), as this would address, in part, some of the concerns about the problems of operationalizing nebulous terms such as ‘worth’, ‘value’, and ‘estimates’.” However, we have chosen not to provide a worked example due to our adding approximately 6 pages addressing the

definitions and limitations of clinical, practical, and statistical significance. Adding more to address this would significantly increase the length of the manuscript. In any case, such a worked example would involve the application of some of the mathematical formulae that we provided, and, as noted by Reviewer 2, “The paper would be better to focus on the meaning of each approach and its pros and cons in a way that all readers can appreciate”.

- (5) The reviewer stated that “step four (p. 20) is probably the heart of the difficulty of measuring economic significance, and, even though it is addressed on pages 21-2, this needs to be addressed in greater detail for the reader”. We agree that Step 4 is at the “heart of the difficulty of measuring economic significance.” As such, we provided more detail, as follows:

In addition, the team should decide for how long it should take to measure each outcome. Further, the data collectors pertaining to each outcome should be identified. Most importantly, the team should determine how to maximize the integrity and fidelity of the data collection process. (p. 27)

- (6) The reviewer stated “it is disingenuous to mention qualitative data (e.g. p. 22) as such data would have to be converted into a metric; further surely it is only ratio level data that could be used (p. 22); to talk of the other three scales is misleading; this is probably just a matter of more careful phrasing in the paper.” With all due respect to the reviewer, we disagree strongly that qualitative data should not be mentioned. While we agree that qualitative data would have to be converted into a metric, this does not stop qualitative data from being collected. For example, mixed methods researchers refer to the concept of “quantitizing,” in which qualitative data are converted into numerical codes that can be represented statistically. As stated by Sandelowski (2001, p. 231), in quantitizing, qualitative data are “numerically represented, in scores, scales, or clusters, in order more fully to describe and/or interpret a target phenomenon.” Also, Boyatzis (1998, p. 129) referred to the counting of themes as “quantitative translation.” Thus, the collection of qualitative data can play an important role in deriving measures of economic significance. For example, researchers could interview participants regarding their perceptions of the utility of the outcome, and then they could quantitize these qualitative data. (For more information about quantitizing data, please see Onwuegbuzie & Teddlie, 2003.) Thus, we have not removed our mention of qualitative data on page 22. Further, we disagree that “it is only ratio level data that could be used.” For example, test score data represent *interval* data. Also, it is not unusual for teachers to rank their students on some outcome. Additionally, as noted previously, nominal (e.g., qualitative) data could be collected. Therefore, we have retained the following question, which now appears on page 27, “What scales of measurement should be used (i.e., nominal, ordinal, interval, ratio)?”
- (7) The reviewer stated that “the paper is marked by clarity and singularity on unclear and complex matters; this needs to be acknowledged, i.e. the practical realities of measuring multi-dimensional phenomena in comparatively simple metrics may misrepresent the complexity of the phenomena under consideration; the paper needs to justify doing this

conceptually as well as practically”. To be honest, we are at a complete loss as to how to address this recommendation. We believe that we have pointed out several limitations of the ESIs, which should caution readers that these indices “may misrepresent the complexity of the phenomena under consideration”. However, isn’t this the case with any index? Even the most complex quantitative techniques such as Structural Equation Modeling (SEM) and Hierarchical Linear Modeling (HLM) “may misrepresent the complexity of the phenomena under consideration.” This does not mean that SEM and HLM cannot play an important role in aiding our understanding of phenomena. And, so it is with ESIs. At the same time, we would like to think that we have “justif[ied] doing this conceptually as well as practically”. Indeed, the reviewer’s statement here appears to contradict his overall assessment of our paper that “the argument about the need for economic significance, coupled with how to calculate it, is well placed and the case is made persuasively.” As such, we did not feel that we were able to address these comments without more explicit directions.

- (8) The reviewer stated that “The ‘summary and conclusion’ section is too long, and repeats earlier material. In fact some of the material could be better placed in the introductory section to the whole paper (suitably downsized), leaving this section only for a very short conclusion, maybe retaining the second main paragraph on page 26 (amongst others) here. The reference to changes in the APA Manual should be excised, though the reference to reporting indications of confidence intervals should be retained.” With all due respect to the reviewer, we disagree that “some of the [summary] material could be better placed in the introductory section to the whole paper.” Because our paper is long, containing 40 pages and discussing several issues, we believe strongly that a summary is justified. Indeed, our summary only contains three paragraphs (1.56%) and the conclusion contains four paragraphs (2.08%), which we believe are reasonable for an article that contains 192 paragraphs. Also, respectfully, we disagree that “the reference to changes in the APA Manual should be excised” because we believe strongly that APA should not only focus on statistical significance and practical significance, but also on economic significance. The APA manual represents a powerful voice of change. Thus, including one or more statements in the APA manual that recommend the use of ESIs when appropriate offers an extremely effective way of promoting ESIs.

## **Reviewer 2**

We are pleased that the reviewer believes, “The author is quite correct in stating that cost is hardly ever stacked up against purported benefit in educational research. For that reason alone, I believe that the paper has a strong message and should be published.”

- (1) We agree with the reviewer’s suggestion to “that some of the mathematics, perhaps all of the section on confidence intervals for the economic tests, is moved to a subsequent appendix. The paper would be better to focus on the meaning of each approach and its pros and cons in a way that all readers can appreciate, leaving us nerds to look at the appendix as and when needed.” We have moved the sections on confidence intervals into an Appendix.

### Reviewer 3

We are glad that the reviewer gleaned that, “This text is well composed, clearly written and carefully structured. From a statistics perspective, the paper is sound and robust. The points made about significance and power, while not new, are exact and well stated.”

- (1) The reviewer suggested, “Clinical significance and practical significance need to be defined and illustrated, including the distinction between them (and with references to the literature).” This is a great suggestion. As stated above, we have added an entire section defining, illustrating, and discussing the limitation of clinical, practical, and statistical significance. Furthermore, we have added a paragraph that discusses the differences between economic significance and the existing three types of significance.
- (2) We agree with the reviewer that “Economic significance needs to be defended as separate from clinical and practical significance. It could be argued that it is a sub-set of clinical significance. This requires close examination.” As noted above, we have included the sections detailed above to clarify these issues.

### References

- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.
- Sandelowski, M. (2001). Real qualitative researchers don't count: The use of numbers in qualitative research. *Research in Nursing & Health*, 24, 230-240.