## Supplemental Materials 1

[Author],

My name is [first author] and I am working (under the direction of [third author]) on a systematic evaluation of what common "gaps" are present in single case design studies that cause these studies to be scored as having low quality or low rigor. In our systematic search, we found your review [title] and would like to include the SCARF rigor/quality data collected for your review in our analysis. All we need is your raw SCARF excel sheet that you used to code for rigor/quality. The goal is to simply take the data from reviews that have used SCARF and use the aggregated data to determine common "gaps" in the rigor/quality of single case designs (e.g. see if researchers are commonly not reporting standardized test scores for their participants).

Best,

[first author]

## Supplemental Materials 2

Changes from SCARF 1.0 to 2.0 (as reported by SCARF authors directly in the excel sheet).

1. We re-arranged categories to more closely align with typical manuscript organization. Categories associated with QUALITY have gray headers, categories associated with RIGOR have black headers and cells are filled with gray; scores associated with OUTCOMES have green headers.
2. We reduced redundancies by removing maintenance and generalization measurement values from the "quality" calculation (since they are considered via graphed maintenance and generalization outcomes), and removed measurement considerations from maintenance and generalization outcomes (e.g., a score of 4 can coded regardless of the time of follow up).
3. We made some minor clarifications to individual categories. Changes are generally indicated in green italicized text.
4. We used the "random" feature in Excel to separate data points that received the same scores (e.g., two designs with a Quality/Rigor score of 2.3 and an Outcome score of 4) by randomly assigning the data point to appear somewhere in the interval range for outcomes (e.g., 4.5 for one design and 4.6 for the other design), given that scores are ordinal and data points anywhere in the 4 range are equal.
5. On maintenance and generalization graphs, we replaced numerical values associated with measurement with descriptive labels.