# Issues and advances in the systematic review of single-case research:

# A commentary on the exemplars

Rumen Manolov, Georgina Guilera, and Antonio Solanas

Department of Social Psychology and Quantitative Psychology,

Faculty of Psychology,

University of Barcelona (Spain)

## Author Note

Correspondence concerning this article should be addressed to Rumen Manolov, Departament de Psicologia Social i Psicologia Quantitativa, Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, 08035-Barcelona, Spain. Phone number: +34934031137. Fax: +34934021359. E-mail: rrumenov13@ub.edu

## Acknowledgement

The authors would like to thank Daniel M. Maggin for his comments on a previous version of the article.

**Abstract**

The current text comments on three systematic reviews published in the special section "Issues and Advances in the Systematic Review of Single-Case Research: An Update and Exemplars". The commentary is provided in relation to the need to combine the assessment of the methodological quality of the studies included in systematic reviews, the assessment of the presence of functional relations via visual analysis following objective rules, and the quantification of the magnitudes of effect, providing meaningful information. Although it was not required that the exemplars follow specific guidelines for conduct and reporting, we applied an existing methodological quality checklist for systematic reviews and meta-analyses. Finally, we point at specific signs of advance in the field of performing systematic reviews of single-case design studies, as identified in the three exemplars, and we also suggest some issues requiring further research and discussion.

*Keywords*: systematic review, meta-analysis, methodological quality, effect size, software

**Meta-Analysis in the Context of Single-Case Research Designs**

Considering that one of the main features of single-case research design s (SCRDs) is their application to one or few participants (Shadish & Sullivan, 2011), establishing the generality of the results requires replication, as statistical generalizations are likely unwarranted (Edgington & Onghena, 2007). Thus, the systematic review of the replications is crucial for assessing whether an intervention can be considered an evidence-based practice (Jenson, Clark, Kircher, & Kristjansson, 2007). One the one hand, it has been questioned whether SCRD data should be meta-analyzed and what interpretations are justified (Strain, Kohler, & Gresham, 1998). More recently, Mark Wolery (personal communication, September 16, 2016) stated that he finds "attempts to do quantitative synthesis of single case studies totally misguided, wrong-headed, and harmful to the field". On the other hand, several special issues have provided arguments for the usefulness of systematic reviews and meta-analyses of SCRD studies: (a) Burns (2012) lists strategies for dealing with threats to the internal validity of meta-analyses and stresses that moderator analysis can provide nuanced information on external validity; (b) Parker and Vannest (2012) highlight that design flexibility, visual analysis, and meaningful effect size indices are complementary; (c) Shadish (2014) underscores that statistical analysis is not restricted to *p* values and mean levels; and (d) the current special issue shows that applied researchers and statisticians can collaborate in performing systematic reviews that are methodologically sound and informative for practitioners looking for evidence regarding intervention effectiveness.

**Tools for Improving Reporting and Methodological Quality**

Researchers willing to conduct and report better systematic reviews and meta-analysis can benefit from consulting the following pieces of information *before* carrying out the research: (a) the exemplars included in the current special issue; (b) the PRISMA statement (Liberati et al.,

2009) on reporting the process and the results; and (c) the AMSTAR checklist (Shea et al., 2007) for assessing the methodological quality[1]. The methodological quality tool was applied independently by the first and second authors to the three exemplars (see Table 1). We briefly mention some aspects, as the authors of the exemplars were not required to follow this checklist.

[1] The AMSTAR checklist has been shown to be a valid and reliable tool (Shea et al., 2009). We used the Scottish Inter-collegiate Guidelines Network (SIGN) checklist, which rephrases the first three AMSTAR items into the first four SIGN items.

**Table 1.** Result of the application of the SIGN checklist to the three exemplars.

| Item | Common et al. | Maggin et al. | Barton et al. |
|---|---|---|---|
| The research question is clearly defined and the inclusion / exclusion criteria must be listed. | Yes | Yes | Yes |
| A comprehensive literature search is carried out. | Yes | Yes | Yes |
| At least two people should have selected studies. | Yes | NES | Yes |
| At least two people should have extracted data. | Yes | Yes | Yes |
| The status of publication was not used as an inclusion criterion. | No | Yes | NES |
| The excluded studies are listed. | NA | Yes | NA |
| The relevant characteristics of the included studies are provided. | Yes | Yes | Yes |
| The scientific quality of the included studies was assessed and reported. | Yes | Yes | Yes |
| Was the scientific quality of the included studies used appropriately? | Yes | Yes | Yes |
| Appropriate methods are used to combine the individual study findings. | Yes | Yes | Yes |
| The likelihood of publication bias was assessed appropriately. | NES | NES | Yes |
| Conflicts of interest are declared. | NA | NA | NA |

*Note.* NA – criterion to be assessed when the final publication is available. NES - Not explicitly stated

In terms of good practice, in Common et al. (this issue) two people selected studies, whereas in terms of transparent reporting, they explicit state that only publications from refereed journals were used. Maggin et al.'s (this issue) review also includes conference papers and technical reports, thus not using publication status as an inclusion criterion. A noteworthy aspect of the Barton et al. (this issue) review is that the possibility of publication bias was explicitly assessed.

The only disagreement in applying the checklist was whether Maggin et al. (this issue) meet the criterion of listing excluded studies: they do list the studies excluded from the meta-analysis, but not the ones excluded from the systematic review, as distinguished in PRISMA statement[2].

**Aims of the Present Commentary**

The commentary on the three exemplars will focus on the following topics: (a) the assessment of functional relations via visual analysis; (b) the effect size indices used; and (c) the assessment of methodological quality. We also point at aspects illustrating the progress in research synthesis and issues needing further research and discussion. Finally, we include the independently provided opinion of an expert in systematic reviews (the second author), who does not work in the SCRD field: each comment ends with a paragraph referred to as the "external perspective".

## Commentary on the Exemplars

**Functional assessment-based interventions for students with or at-risk for high incidence disabilities: field-testing single-case syntheses** (Common, Lane, Pustejovsky, Johnson, & Johl). First, the visual analysis used for assessing the presence of a functional relation is based on the criteria by Gast and Spriggs (2014), which enhance the objective and systematic application of visual analysis. Additional assessment of the difference between conditions could be performed

[2] Following the PRISMA statement (Liberati et al., 2009, p. 1), "The conduct of a systematic review comprises several explicit and reproducible steps, such as identifying all likely relevant records, selecting eligible studies, assessing the risk of bias, extracting data, qualitative synthesis of the included studies, and possibly meta-analyses. (…) If quantitative synthesis is performed, this last stage alone is referred to as a meta-analysis."

using the conservative dual criterion for stable data (Fisher, Lomas, & Kelley, 2003) or projecting trends considering data variability (Manolov, Sierra, Solanas, & Botella, 2014).

Second, regarding the effect sizes used, this exemplar illustrates well the challenges that single-case meta-analysts face. The between-cases standardized mean difference (BC-SMD) is a reasonable choice, but it is applicable to few of the studies included in the systematic review. In order to be able to compute an effect size for each participant, the log response ratio is used, which is also justified, as it yields meaningful information in terms of percentage change and also allows obtaining confidence intervals around the summary measures. Further challenges faced in the use of the log response ratio are the assumptions of stable data and lack of autocorrelation, the second of which is dealt with by using robust variance estimation.

Third, in terms of assessing the methodological quality of the studies included in the systematic review, Common et al. (this issue) use a weighted coding scheme that attenuates the requirement to meet all 22 components of all eight quality indicators of the Council for Exceptional Children *Standards* (CEC, 2014), although the main evaluation is performed using the original criteria and the attenuated ones are used for descriptive purposes only. Regarding the assessment of methodological quality, in our perspective, the application of the CEC *Standards* appears to provide further restrictions to the What Works Clearinghouse (WWC) *Standards* (Kratochwill et al., 2010): according to WWC, a study demonstrating an intervention effect three times at three different points in time meets evidence standards, but according to CEC, a positive effect of the intervention requires at least three participants per study. Accordingly, differences between rubrics have already been identified (Maggin, Briesch, Chafouleas, Ferguson, & Clark, 2014). We consider that in absence of consensus or convergence between rubrics, it would be wise for research synthesists to either apply the rubric most commonly used in their field or

apply more than one rubric and report how the results of the evaluation of methodological quality (e.g., percentage of studies meeting criteria) agree or differ according to the rubric.

Regarding the signs of advance in the field, Common et al. (this issue) show how, in the context of the same systematic review, information can be provided about: (a) whether the studies included are methodologically sound; (b) whether there is enough evidence for a positive effect of the intervention, and (c) about the magnitude of the effect of the intervention, considering potential moderators. The results of the systematic review can be used not only for labeling or not Functional assessment-based interventions as an evidence-based practice, but also for further emphasizing the need to carry out methodologically sound studies.

Concerning the aspects still in need of development, it is important that researchers be aware of (and deal with) the dependency between outcomes belonging to the same study: Common et al. use robust variance estimation and an explicit mention that this procedure is also useful for that purpose would have addressed this topic. A second aspect that requires further clarification in order to guide researchers performing systematic reviews is whether the publication bias has to be evaluated statistically or it can be gauged according to the degree of transparency and comprehensiveness of the literature search. A third aspect requiring more space is a discussion of generalization to other individuals in order to make specific recommendations for practice.

Finally, the external perspective highlights authors' efforts dedicated to identifying relevant records in diverse sources (i.e., four electronic databases, ancestral searchers, hand searchers in six journals, and contact with two authors), meeting all criteria for a comprehensive search (Higgins et al., 2013). An additional strength refers to the thoroughness of the coding process for guaranteeing its reliability (i.e., training of coders and at least 85% of inter-rater agreement). Transparency and replicability may benefit from either listing the studies that did not fulfill the

inclusion criteria (e.g., in a supplementary material), describing the reasons for their exclusion (e.g., in a flow diagram), or both. Such information, especially in relation to studies excluded due to lack of information or poor methodological quality, would help enhancing future research.

**A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update** (Maggin, Johnson, Pustejovsky, Shadish, & Chafouleas)

First, visual analysis for assessing the presence of a functional relation is applied similarly to Common et al. (this issue) and, additionally, the results of the different quantifications involved are provided, which is recommendable, as it makes the process more transparent.

Second, regarding the effect sizes used, Maggin et al. (this issue) used two different models for the BC-SMD: one assuming an immediate and stable effect and the other one allowing for a progressive effect[3]; due to the similarity of their results, the authors advocate for using of the more parsimonious one. In terms of reporting, two aspects are noteworthy: (a) the detail provided in Maggin et al.'s Table 2S and (b) the comments on the intention to include further moderators, which could not be materialized due to of lack of variation and of information. In relation to moderator analysis, we consider that a meta-analysis could be strengthened by specifying the theoretical or empirical bases for selecting moderators and also by discussing not only the results of statistically significant moderators (when such are found), but also by further examining of the negative results of the moderator analysis. Specifically, researchers could consider whether negative results are related to the number of studies included (i.e., low statistical power) or to the (lack of) availability of information on participant, intervention, or setting characteristics.

[3] Transparency and replicability would benefit from using a statistical model available elsewhere apart from an unpublished doctoral dissertation (e.g., for the gradual effect model, the exact expression and the number of measurements and number of cases required are not specified).

Third, for assessing the methodological quality of the studies, Maggin et al. (this issue) use the WWC *Standards* (Kratochwill et al., 2010) and code the information about further ten methodologically relevant items, an excellent attempt to evaluate the rigor of the studies.

Regarding the signs of advance in the field, this exemplar shows that it is possible, using only studies meeting *design* standards, to provide a statistical quantification of effect size jointly with a tally of the studies meeting *evidence* standards. In terms of the conclusions, specifying the conditions mostly represented by the studies reviewed favors the assessment of generalizability.

Concerning the aspects still needing development, an important aspect arises in terms of how to compare the results reported in an update and the results reported in the original review. For instance, the exemplar by Maggin et al. (this issue) updates a study by Maggin, Johnson, Chafouleas, Ruberto, and Berggren (2012), using five different quantifications of effect. On the one hand, comparability between effect sizes is aided by the fact that three of the effect size measures used by Maggin et al. (2012) were expressed in standard deviations. On the other hand, in terms of convergence or divergence of findings, for instance, in Maggin et al. (2012) for all standardized measures values greater than the cut-off point of 2.0 were obtained, in contrast with both BC-SMD values from the update, which are lower than 2.0. In that sense, developing reasonable guidelines and cut-off points would be beneficial for future research.

The external perspective applauds the emphasis that Maggin et al. (this issue) on making visual analysis not only more transparent, but also objectively replicable. Another noteworthy contribution is that the authors present their results separately (i.e., previous review vs. updated review) and conjointly, allowing an examination of trends in SCRD. To find that none of the moderators (out of nine related to sample, intervention, and methodological characteristics)

explain heterogeneity in effect sizes deserves an attempt to speculate on why and what other relevant factors may contribute to the explanation of variation in intervention effects.

**A meta-analysis of technology aided instruction and intervention for students with ASD** (Barton, Pustejovsky, Maggin, & Reichow, this issue)

First, visual analysis is used for assessing the presence of a functional relation using the criteria from the WWC *Standards* (Kratochwill et al., 2010) and from Gast and Spriggs (2014), but more detail about the process and its results could have been provided (e.g., Maggin et al., this issue).

Second, regarding the effect sizes used, using the BC-SMD made possible comparing the results provided by group-comparison and single-case studies. In terms of transparency, the authors mention in several places of the text that in order to apply the BC-SMD more than half of the single-case studies were excluded from the meta-analysis. In terms of interpretation of the results, a distinction was made between different types of intervention and between the skills taught. Once heterogeneity has been identified, a meta-analysis would benefit from a moderator analysis or from suggestions of potentially relevant moderators.

Third, in terms of assessing the methodological quality of the studies, Barton et al. (this issue) advocate for instruments assessing specific threats to internal validity instead of more general quantifications of methodological rigor. Here arises a relevant issue: the importance of justifying the choice of using a specific tool for assessing risk of bias, as other *published* scales (e.g., Tate et al., 2013) serve the same purpose.

Regarding the signs of advance in the field, this exemplar takes full benefit of the advantages of the BC-SMD by integrating the results of group design and SCRD and assessing publication bias. Moreover, the overall summary of the evidence via the 5-3-20 rule (Kratochwill et al., 2013) combines an evaluation of design standards, risk of bias, and presence of a functional

relation as identified via visual analysis. Finally, reporting the summary measure both with and without some effect sizes (estimated with low precision, potential outliers) favors transparency.

Concerning the aspects still needing development, we echo Barton et al.'s (this issue) comment that the nature of publication bias is likely to be related to the assessment of functional relations and refer the interested reader to Shadish, Zelinsky, Vevea, and Kratochwill (2016). In our view, a logical aim would be to include only studies meeting *design* standards, but not necessarily presenting a specific data pattern providing moderate or strong evidence for a functional relation (Kratochwill et al., 2010). A second aspect requiring further attention refers to the role of statistical analysis in the 5-3-20 rule used by Barton et al. For instance, Maggin et al. (2012) establish a cut-off point of 2.0 standard deviations for labeling an intervention effective, but it is not clear whether such a criterion is additional to or an alternative for visual analysis.

The external perspective underscores that Barton et al. (this issue) set out the problem of missing information in original studies when reporting relevant data (e.g., sample characteristics), that in turn may have an undesirable effect on the analysis of moderators. Additionally, being aware that most common procedures for assessing publication bias (e.g., Egger's regression test) have not been tailored to SCRD research, the authors have desirably decided to explore outcome reporting bias acknowledging that results should be interpreted with caution. Practitioners may benefit from having a clearer insight to settings, populations and conditions the intervention might generalize.

## Discussion

On the basis of the exemplars and general methodological knowledge, it is possible to identify five pillars of the assessment of the evidence-base of interventions studied via SCRD designs. First, at the within-study level, SCRDs are flexible and rigorous enough to provide causal

evidence regarding the effectiveness of interventions for specific clients. Second, assessing the methodological quality of studies makes possible reviewing systematically only studies capable of providing strong evidence. Third, the presence of a functional relation can be assessed via visual analysis that is as transparent and structured as possible (Fisher et al., 2003; Gast & Spriggs, 2014; Kratochwill et al., 2010; Lane & Gast, 2014), as recent evidence suggests that training in structured criteria is beneficial (Wolfe & Slocum, 2015; Young & Daly III, 2016). Fourth, the exemplars illustrate summary measures for which confidence intervals can be constructed, heterogeneity can be quantified, and the presence of publication bias can be assessed, which is likely to make the evidence from SCRDs credible (Shadish, Rindskopf, & Hedges, 2008). Fifth, Maggin et al. (this issue) illustrate that systematic reviews can include both a meta-analytical summary measure and a tally of the number of studies meeting evidence standards are met. These two pieces of information can lead to a qualitative judgment on whether an intervention is an evidence based practice (e.g., Barton et al., this issue).

Regarding topics for future research and discussion, in the systematic review by Maggin et al. (this issue), it is suggested that methods with less stringent data requirements have to be developed. Accordingly, we point at two within-case indices not restricted to reversal or multiple baseline designs. First, the Nonoverlap of all pairs (NAP; Parker & Vannest, 2009) is computationally equivalent to an indicator called probability of superiority, for which Grissom and Kim (2001) provided an expression for estimating its variance[4]. Second, for the percentage change index (PCI; or mean baseline reduction; Olive & Smith, 2005) an expression for the

[4] According to page 141 of Grissom and Kim (2001): $Var = (1/12)[(1/m) + (1/n) + (1/mn)]$, where *n* and *m* are the sizes of the baseline and intervention phase, respectively.

variance[5] was provided by Hershberger, Wallace, Green, & Marquis (1999). For both NAP and PCI the expressions for estimating variance (necessary for using inverse variance weighting) are based on the assumption of independent data and, therefore, more research is necessary on the validity of these expressions, given that single-case data that are likely to be serially related (Shadish & Sullivan, 2011). A third analytical option are multilevel models (Moeyaert, Ferron, Beretvas, & Van Den Noortgate, 2014), because they allow (a) taking into account the dependencies arising from measurements being nested into participants and participants (and their outcomes) into studies; (b) modeling several data features at the within-study level (e.g., trend, autocorrelation); and (c) incorporating moderator analysis at the across studies level.

Finally, note that William Shadish appears in the current commentary as one of the driving forces behind several special issues (e.g., Shadish et al., 2008), as an author of an extensive review of the main features of published single-case research (Shadish & Sullivan, 2011), and also as a co-author of one of the most promising statistical developments in the single-case designs context (Shadish et al., 2014). In his honor, we should spread the idea that statistics is more than just *p* values and help applied researchers to obtain the quantifications with the appropriate user-friendly software[6] as he suggested (Shadish, 2014), while also helping them interpret the results correctly and in a meaningful way.

---

[5] According to page 122 of Hershberger et al. (1999): $Var = 1/s_A^2 \left(\frac{s_A^2+s_B^2}{3} + \frac{(\bar{Y}_A-\bar{Y}_B)^2}{2}\right)$, where $s_A^2$ is the variance of the baseline data, $s_A^2$ is the variance of the intervention phase data, $\bar{Y}_A$ is baseline mean and $\bar{Y}_B$ is the intervention phase mean; always referring to the last three measurements per phase.

[6] At https://osf.io/t6ws6/, we are offering a list of software tools (web-based applications, code in R, SPSS, and SAS, and Microsoft Excel macros).

## References

Barton, E. E., Pustejovsky, J. E., Maggin, D. M., & Reichow, B. (2017). A meta-analysis of technology aided instruction and intervention for students with ASD. *Remedial and Special Education*.

Burns, M. K., (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education, 21*, 175-184.

Common, E. A., Lane, K. L., Pustejovsky, J. E., Johnson, A. H., & Johl, L. E. (2017). Functional assessment-based interventions for students with or at-risk for high incidence disabilities: field-testing single-case syntheses. *Remedial and Special Education*.

Council for Exceptional Children (CEC). (2014). *CEC standards for evidence-based practices in special education*. Arlington, VA: Author.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387-406.

Gast, D. L., & Spriggs, A. D. (2014). Visual analysis of graphic data. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.) (pp. 176-210). London, UK: Routledge.

Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods, 6*, 135–146.

Hershberger, S. L., Wallace, D. D., Green, S. B., & Marquis, J. G. (1999). Meta-analysis of single-case data. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 107-132). London, UK: Sage.

Higgins, J., Lane, P. W., Anagnostelis, B., Anzures-Cabrera, J., Baker, N. F., Cappelleri, J. C., ... & Whitehead, A. (2013). A tool to assess the quality of a meta-analysis. *Research Synthesis Methods, 4*, 351-366.

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools, 44*, 483-493.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*, 445-463.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine, 6*(7), e1000100.

Maggin, D. M., Briesch, A. M., Chafouleas, S. M., Ferguson, T. D., & Clark, C. (2014). A comparison of rubrics for identifying empirically supported practices with single-case research. *Journal of Behavioral Education, 23*, 287–311.

Maggin, D. M., & Chafouleas, S. M. (2013). Introduction to the Special Series: Issues and advance of synthesizing single-case research. *Remedial and Special Education, 34*, 3-8.

Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology, 50*, 625-654.

Maggin, D. M., Johnson, A. H., Pustejovsky, J. E., Shadish, W. R., & Chafouleas, S. M. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education*.

Manolov, R., Sierra, V., Solanas, A., & Botella, J. (2014). Assessing functional relations in single-case designs: Quantitative proposals in the context of the evidence-based movement. *Behavior Modification, 38*, 878-913.

Moeyaert, M., Ferron, J., Beretvas, S., & Van Den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of since-case experimental designs. *Journal of School Psychology, 52*, 191-211.

Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology, 25*, 313-324.

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357-367.

Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education, 21*, 254-265.

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology, 52*, 109-122.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*, 123-147.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 188-196.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43*, 971-980.

Shadish, W. R., Zelinsky, N. A. M., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49*, 1–18.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 7*, 1.

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., ... & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology, 62*, 1013-1020.

Strain, P. S., Kohler, F. W., & Gresham, F. (1998). Problems in logic and interpretation with quantitative syntheses of single-case research: Mathur and colleagues (1998) as a case in point. *Behavioral Disorders, 24*, 74-85.

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakima, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation, 23*, 619-638.

Wolfe, K., & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of AB graphs. *Journal of Applied Behavior Analysis, 48*, 472–477.

Young, N. D., & Daly III, E. J. (2016). An evaluation of prompting and reinforcement for training visual analysis skills. *Journal of Behavioral Education, 25*, 95-119.