Romanian Journal of School Psychology, December 2025
Vol. 18, No. 35-36, pp. 53-64

DATA ANALYSIS IN SINGLE-CASE EXPERIMENTAL DESIGNS (SCED FOR EDUCATIONAL INTERVENTIONS IN AUTISM SPECTRUM DISORDER: STATISTICAL RECOMMENDATIONS

Angelica Hălmăjan [1]

University of Oradea
Romania

Abstract
Single-case experimental designs (SCED) occupy a central role in applied educational psychology for autism spectrum disorder (ASD), particularly in behavior-focused interventions (e.g., reducing challenging behaviors, increasing learning responses) and in augmentative and alternative communication (AAC) approaches such as the Picture Exchange Communication System (PECS). However, the utility of SCED critically depends on the quality of causal inference and the appropriateness of data-analytic strategies. This article synthesizes recommended statistical approaches for SCED, with a focus on (a) integrating visual analysis with quantitative effect estimation, (b) controlling for trend and serial dependence, (c) segmented regression models and randomization tests, and (d) standardized effect sizes suitable for meta-analyses and comparison with between-group studies. The advantages and limitations of non-overlap indices (e.g., Tau-U), between-case standardized mean differences (BC-SMD), and response ratio effects (log response ratio) are examined. Current standards for design and reporting (WWC v5.0; SCRIBE 2016) are reviewed, and a practical decision-making framework for selecting analytic strategies based on research questions, outcome characteristics, and phase structure is proposed. Finally, concrete recommendations are provided for researchers and practitioners in special education implementing ABA- and PECS-based interventions for ASD.

Keywords: single-case experimental designs (SCED); single-subject design; autism spectrum disorder (ASD); applied behavior analysis (ABA); PECS; visual analysis; Tau-U; BC-SMD; randomization tests; segmented regression

[1] Lecturer, PhD, PhD, Department of Psychology, Faculty of Social and Humanistic Sciences, University of Oradea, Romania; E-mail: abrihan@uoradea.ro

In educational psychology, interventions for ASD are frequently evaluated using single-case experimental designs (SCED) because (i) populations are heterogeneous, (ii) interventions are highly individualized, (iii) outcomes are operationally defined and repeatedly measured, and (iv) ethical and equity considerations may constrain group-level randomization. Importantly, SCED should not be regarded as a "weak design", but rather as a family of rigorous experimental designs characterized by repeated measurement, systematic manipulation of the independent variable, and replicated demonstration of effects. Evidence-based practice initiatives in ASD indicate that a substantial proportion of the empirical literature supporting the efficacy of focused interventions, including behavioral interventions and AAC/PECS-based approaches, derives from SCED studies, underscoring the importance of methodological quality for translation into practice (Steinbrenner et al., 2020).

In ABA-based interventions, behavioral targets commonly include response rates or proportions (e.g., frequency of functional requests, percentage of tasks completed, frequency of stereotyped behavior), measured through direct observation and aggregated across sessions. In PECS interventions, progress is frequently assessed through the acquisition of exchanges, as well as generalization and maintenance across school and family settings (Bondy & Frost, 2001; Frost & Bondy, 2002). Analyses of such outcomes must therefore address serial dependence, trend, intra-phase variability, and potential gradual or delayed intervention effects (e.g., slow increases, acquisition thresholds, saturation effects).

1. Design requirements and implications for analysis

The selection of statistical analyses cannot be dissociated from design quality. Contemporary standards for SCED used in evidence syntheses require clearly defined phases, an adequate number of measurement points per phase, replication of effects, and explicit control of threats to internal validity. For example, the What Works Clearinghouse (WWC) specifies criteria for evaluating SCED within its educational evidence standards (What Works Clearinghouse, 2022). Complementary recommendations further emphasize transparency in randomization and counterbalancing procedures (when applicable), systematic reporting of intervention fidelity, pre-specification of analytic strategies, and explicit handling of trend and serial dependence (Kratochwill et al., 2023).

From an analytic perspective, alignment with design characteristics is essential. In multiple-baseline designs, replicated effects across participants, behaviors, or settings constitute the core evidential criterion. In ABAB or withdrawal designs, reversibility and potential carryover effects must be explicitly considered. In changing-criterion designs, the adequacy of criterion steps and their temporal separation directly support causal inference. Across designs, the most robust analytic approaches preserve

temporal structure, explicitly quantify changes in level and/or slope, provide estimates of uncertainty (e.g., confidence intervals or standard errors), and allow replication and aggregation across cases.

2. Visual analysis: importance, criteria, and operationalization

Visual analysis remains a core component of single-case experimental designs (SCED), particularly in ASD research, as it supports clinical and educational judgements regarding the functional significance of change. To be informative, visual analysis must be operationalized through systematic evaluation of level, trend, variability, immediacy of effect, overlap, and consistency of effects across replications. In advanced research contexts, visual analysis is not treated as a substitute for statistical inference but as a complementary analytic component that informs model selection, highlights potential artefacts (e.g., outliers or contextual changes), and guides interpretation.

Current methodological guidance recommends a dual-criteria approach: (a) standardized graphical presentation of data, including consistent axes, clear phase demarcation, and reporting of measurement units and session conditions; and (b) quantitative estimation of intervention effects, accompanied by explicit consideration of trend and serial dependence. The SCRIBE 2016 guidelines provide a structured framework for reporting design, measurement, intervention fidelity, and analysis in a manner that enables replication and critical appraisal (Tate et al., 2016).

3. Non-overlap indices: utility and risks

Non-overlap indices are effect size measures primarily used in single-case experimental designs (SCED; e.g., AB, ABA, ABAB, multiple-baseline designs) to quantify the extent to which values observed during the intervention phase (B) are separated from those observed during the baseline phase (A). The underlying rationale is that, if an intervention is effective, data points in phase B should systematically improve upon those in phase A, resulting in minimal overlap between the two distributions.

Non-overlap indices are most appropriate under the following conditions:

a) When samples are small or time series are short, making classical parametric statistics (e.g., t-tests) inappropriate or unreliable;
b) When a rapid and easily communicable quantification of the difference between phases a and b is required, often expressed as percentages or probabilities;
c) In SCED meta-analyses, where such indices allow standardization of effect reporting across studies;
d) When outcome variables have a clearly defined direction of improvement (e.g., “higher values indicate improvement” or “lower

values indicate improvement"), such that the meaning of change is unambiguous.

However, non-overlap indices have important methodological limitations. They do not account for baseline trend or serial dependence and may therefore misestimate effects when outcomes change over time. They are sensitive to ceiling and floor effects and to extreme values, particularly in short series. Tied observations, common with categorical or coarse measures, are handled inconsistently across indices. Importantly, non-overlap indices do not replace visual analysis and may appear favorable even when change is gradual, delayed, or highly variable. Differences in phase length and measurement procedures may further limit comparability across studies.

The main indices of non-overlap are:

3.1 Percentage of Non-Overlapping Data (PND) measures the proportion of intervention-phase data points (B) that exceed (or fall below, depending on the desired direction of change) the most extreme data point observed in the baseline phase (A). PND is appropriate when the baseline is relatively stable and a simple, rapidly interpretable indicator of intervention effectiveness is required.

3.2 However, PND has important limitations. It is extremely sensitive to a single extreme baseline value, such that an outlier can yield a very low PND even when phase B shows clear improvement. The index strongly penalizes baseline variability and does not reflect the magnitude of change, indicating only whether a threshold is exceeded. For these reasons, PND is generally considered one of the most methodologically fragile non-overlap indices.

3.3 Percentage of Data Exceeding the Median (PEM) represents the proportion of intervention-phase data points (B) that exceed the median value of the baseline phase (A). PEM is best suited to baseline data containing outliers, as the median provides greater robustness than extreme-value–based indices such as PND.

3.4 PEM has notable limitations. It does not adjust for baseline trend, and the presence of systematic change in phase A may lead to overestimation of intervention effects. In addition, when the distributions of phases A and B overlap substantially, PEM may yield intermediate values that are difficult to interpret in clinically meaningful terms.

3.5 Percentage of All Non-Overlapping Data (PAND) represents the proportion of observations remaining after removing the minimum number of data points required to eliminate all overlap between baseline (A) and intervention (B) phases. PAND is useful when an approximate estimate of overall phase separability is desired.

3.6 PAND has important limitations. The index is influenced by phase length, with longer phases potentially producing greater overlap, and it

can be more difficult to compute and explain than simpler non-overlap measures. In addition, PAND may obscure temporal dynamics, such as gradual versus abrupt change

3.7 Improvement Rate Difference (IRD) measures the difference between improvement rates in the intervention (B) and baseline (A) phases and is conceptually related to risk or rate difference measures. IRD is useful when an interpretable indicator, such as a "success rate difference" between phases, is required.

3.8 IRD has limitations. The index may be unstable in short time series or in the presence of numerous tied values and does not explicitly account for baseline trend or serial dependence.

3.9 Nonoverlap of All Pairs (NAP) compares all baseline–intervention data pairs and estimates the proportion of pairs in which the intervention value exceeds the baseline value, with tied pairs typically weighted as 0.5. NAP provides an intuitive interpretation as the probability that a randomly selected data point from phase B is better than one selected from phase A.

3.10 NAP is useful when a more robust alternative to PND or PEM is desired, as it incorporates all pairwise information rather than relying on a single threshold. Its limitations include susceptibility to baseline trend and serial dependence; moreover, gradual improvement with substantial early overlap may yield modest NAP values despite clinically meaningful effects. For these reasons, NAP is often preferred as a pragmatic balance between simplicity and robustness.

Practical recommendations - to reduce risks:

- Combine non-overlap indices with visual analysis, explicitly evaluating level, trend, variability, immediacy of effect, latency, and consistency across replications.
- When baseline trend is present, include a trend-adjusted index, such as Tau-U, which is specifically designed to control for monotonic change.
- Report uncertainty estimates whenever possible (e.g., confidence intervals or bootstrap intervals) and explicitly document phase lengths.
- Specify a priori the direction of improvement (increase vs. decrease) and clearly describe how tied values are handled.
- Assess sensitivity to ceiling/floor effects and extreme values, for example by conducting analyses with and without influential outliers.

Many non-overlap indices are sensitive to outliers, baseline trend, and asymmetric distributions. A more robust and widely used alternative is Tau-U, which integrates information on non-overlap between phases with trend estimation and allows correction for unwanted baseline trend (Parker et al., 2011).

Baseline-corrected Tau (Tau-BC) is a non-parametric effect size index for single-case designs (e.g., A–B or interrupted time-series designs)

that estimates the phase contrast between baseline (A) and intervention (B) after adjusting for baseline trend. This adjustment reduces the risk of inflating intervention effects when phase A already exhibits systematic change.

Tau-BC is obtained by first assessing baseline trend using Kendall's Tau and, when relevant, detrending the data using a non-parametric slope estimator such as the Theil–Sen estimator. The effect is then computed as a tau-type association between phase membership (A = 0, B = 1) and the original or detrended outcome values.

Tau-BC yields values ranging from −1 to +1. Positive values indicate higher outcome levels in phase B than in phase A when higher values represent improvement, whereas negative values indicate lower outcome levels in phase B when reduction is the desired outcome. Interpretation therefore depends on the predefined direction of improvement.

The difference from Tau-U

Tau-U combines non-overlap with information on trend and allows adjustment for baseline trend, but uses a different formulation. In common implementations (e.g., SingleCaseES in R), Tau-U is described as an extension of Kendall's Tau with correction for trend in phase A. By contrast, baseline-corrected Tau (Tau-BC; Tarlow) controls baseline trend through a two-step approach in which trend is assessed and, if necessary, removed prior to estimating the phase contrast. Tau-U (or baseline-corrected variants) is useful for short series, rate or proportion outcomes, and situations in which baseline trend or gradual effects are suspected. Limitations include: rank-based interpretation may be less intuitive for applied decisions, and confidence intervals may be biased when serial dependence is substantial. Accordingly, Tau-U is best used as a complementary indicator rather than as sole evidence of intervention effects.

4. Segmented regression models and models with correlated errors

Segmented regression models (piecewise or interrupted time-series variants in SCED) explicitly estimate changes in level and slope at phase transitions. Specifically, these models quantify both an immediate level change following intervention onset and a change in the rate of change over time. They are particularly suitable for ABA- and PECS-based interventions in which effects are gradual (e.g., skill acquisition) or evolve through adaptation and planned generalizations. A central methodological challenge is serial dependence, as consecutive observations are typically correlated. Ignoring autocorrelation may lead to underestimated standard errors and inflated Type I error rates. Recommended approaches include explicit modelling of autocorrelation, the use of robust standard errors in appropriate settings, or Bayesian models that directly specify temporal dependence. Current software implementations, such as the scan package in R, provide

segmented regression models, multilevel extensions, and Bayesian options specifically tailored to SCED data (Wilbert & Lueke, 2025).

5. Randomization tests and exact inference

When a single-case design includes a genuine randomization component (e.g., randomized intervention onset in multiple-baseline designs or randomized condition order in alternating-treatments designs), randomization tests provide statistical inference under minimal distributional assumptions. These tests evaluate how rare the observed effect is relative to all allocations permitted by the design.

In educational research, where time series are often short and outcome distributions may deviate from normality, randomization tests are particularly attractive alternatives to parametric methods. Their primary limitation is that validity depends on true a priori randomization and strict adherence to the allocation scheme. Consequently, randomization procedures, the randomization space, and the tested statistics must be explicitly documented in the method section and retained as part of the study record, in accordance with SCRIBE recommendations (Tate et al., 2016).

6. Comparable standardized effects between studies: BC-SMD and meta-analysis

For integrating SCED evidence into syntheses and guidelines, it is useful to estimate effect sizes comparable to d from between-group studies. A key benchmark is the standardized mean difference for SCED, which incorporates procedures for estimating autocorrelation and between-case variability and is suitable for meta-analytic applications (Shadish et al., 2014). More recently, between-case standardized mean difference (BC-SMD) methods have been developed, treating measurements as nested within cases and providing flexible estimates across designs and data structures (Chen et al., 2023).In practice, BC-SMD is recommended when at least a small number of cases are available (≥3–4) and when the aim is to report standardized effects with confidence intervals and facilitate comparison with group studies or future meta-analyses. Software such as *scdhlm* and *scan* provide implementations for hierarchical SCED models and BC-SMD estimation (Pustejovsky et al., 2014; Wilbert & Lueke, 2025).

7. Response ratio effects: log response ratio for behavioral data

For behavioral variables collected as rates or proportions (very common in ABA/PECS), an intuitive measure is the proportional change from baseline. For this purpose, log response ratio (lnRR) allows quantification of the effect as relative multiplication/decrease (e.g., "40% decrease in the frequency of the target behavior"), making it suitable for

meta-analytic aggregation. Pustejovsky (2018) describes methods for estimating lnRR in SCED and combining them in meta-analyses, paying attention to the variance issue when there is autocorrelation. Practical recommendation: lnRR is especially useful when the outcome has a significant zero (absence of behavior) and when relative interpretation is preferred by practitioners (e. g., percentage reduction of interfering behaviors). However, careful definition of the unit of measurement and the time window (session, day, activity block) is necessary to avoid misleading comparisons between cases or contexts.

8. Choosing the analysis based on the question and structure: a decision-making guide

In practice, the choice of analysis can be guided by three questions: (1) What form of change is plausible (immediate jump vs. gradual trend)? (2) What is the purpose of inference (educational decision for a case vs. generalizations/synthesis)? (3) What constraints does the design impose (replication, randomization, number of points and cases)?

Table 1 summarizes operational recommendations.

Table 1. Operational recommendations

Situation / question	Recommended primary analysis	Useful secondary indicator
Decision for a single student; AB/ABA/ABAB; possible immediate effect	Segmented regression (level + slope) with correlated errors (AR/GLS)	Tau-U
Gradual acquisition (ABA/PECS); multiple baselines on behaviors / settings	Segmented regression per case + synthesis between cases (hierarchical model)	Tau-U / NAP
	Randomization (permutation) test using pre-specified test statistics	
Design with explicit randomization (moment/order of conditions)	BC-SMD (within cases) with confidence intervals	Effect estimation (BCSMD or lnRR) lnRR for rates/proportions
Synthesis and comparability with group studies; ≥3–4 cases		

Methodological notes

- Report autocorrelation and check the trend in the dataset.
- Emphasis on replication between levels and on maintenance / generalization.
- Validity depends on actual randomization and the allocation space.

– Requires detailed reporting of data and the structure of phases.

9. Reporting, transparency, and reproducibility

To achieve a high standard of publishability, SCED reporting must be complete and reproducible. SCRIBE 2016 recommends reporting the following elements: (a) justification of the design, (b) detailed description of the participant and context, (c) operationalization of variables and measurement procedures, (d) fidelity assessment of the intervention and control procedures, (e) an explicit analysis plan (visual and statistical), and (f) access to sufficiently detailed data or figures to allow reanalysis (Tate et al., 2016).

In educational research, the WWC v5.0 standards emphasize minimum criteria regarding the number of measurement points per phase and the replication of effects across phases or cases, with direct implications for analytic power and the credibility of causal inference (What Works Clearinghouse, 2022).

To support transparency and synthesis, researchers should publish graphs in formats that allow numerical extraction or provide raw data in an appendix, report preprocessing decisions (e.g., handling of missing data and aggregation rules), and include both visual analysis and at least one quantitative indicator with uncertainty estimates (e.g., confidence intervals or standard errors). When software packages are used (e.g., R, Python, SPSS), the software version and relevant settings for autocorrelation estimation and trend correction should be reported.

10. Methodological example: analysis plan for a PECS intervention on functional requests

To illustrate the alignment between the educational question, design, and analysis, we consider a PECS intervention aimed at increasing functional requests (exchanges) in a student with ASD. A recommended design is a multiple-baseline across settings (classroom, office, home) or across behaviors. Outcomes are measured as the rate of exchanges per session or the proportion of opportunities utilized, with systematic recording, inter-observer agreement, implementation fidelity, and data on generalization.

The proposed analysis includes graphical display with phase marking, segmented regression to estimate level and slope change, estimation of autocorrelation with correlated errors when appropriate, Tau-U as a trend-adjusted indicator, and synthesis across settings or cases using hierarchical models or BC-SMD when sufficient cases are available (Chen et al., 2023; Parker et al., 2011). For practitioner-oriented reporting, lnRR may also be reported (Pustejovsky, 2018). A strong effect would be indicated by synchronized level change at intervention onset, sustained

improvement during intervention, reduced overlap between phases, maintenance, and generalization across contexts.

11. Synthetic recommendations for publications

Pre-register, when possible, criteria for phase changes and primary statistics (especially when using randomization tests).Explicitly report serial dependence and the method used to address it (e.g., AR/GLS, robust estimators, Bayesian modelling).Combine visual analysis with at least one quantitative effect estimate (e.g., Tau-U, BC-SMD, lnRR) and report associated uncertainty (CI or SE).Ensure replication of effects across series, behaviors, or cases, and discuss generalization and maintenance as educational criteria. Align reporting with SCRIBE 2016 and, when targeting evidence syntheses in education, address compatibility with WWC standards.

Conclusions

In SCED applied to ASD, recommended analysis does not rely on a single statistical test, but on a coherent set of methodological decisions: (a) a design sufficiently strong to support causal inference, (b) standardized visual analysis, (c) quantitative modelling that separates level and trend and addresses serial dependence, and (d) transparent reporting that enables replication and synthesis. For interventions targeting behavioral outcomes in ABA and communication outcomes in PECS, the combined use of segmented regression models with correlated errors, Tau-U as a robust indicator, and standardized effect sizes (BC-SMD or lnRR) provides a balance between practical utility and methodological rigor. In line with current standards, improvements in analytic quality must be accompanied by increased transparency in reporting and by clear documentation of intervention implementation (Kratochwill et al., 2023; Tate et al., 2016; What Works Clearinghouse, 2022).

References

Bondy, A., & Frost, L. (2001). The picture exchange communication system. *Behavior Modification*, *25*(5), 725–744. https://doi.org/10.1177/0145445501255004

Chen, M., Pustejovsky, J. E., Klingbeil, D. A., & Van Norman, E. R. (2023). Between-case standardized mean differences: Flexible methods for single-case designs. *Journal of School Psychology*, *98*, 16–38. https://doi.org/10.1016/j.jsp.2023.02.002

Frost, L., & Bondy, A. (2002). *The picture exchange communication system training manual* (2nd ed.). Pyramid Educational Products.

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., & Johnson, A. (2023). Single-case intervention research design standards: Additional proposed upgrades and future directions. *Journal of School Psychology, 97,* 192–216. https://doi.org/10.1016/j.jsp.2022.12.002

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, *68*, 99–112. https://doi.org/10.1016/j.jsp.2018.02.003

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, *39*(4), 211–227. https://doi.org/10.3102/1076998614547577

Pustejovsky, J., Chen, M., & Hamilton, B. (2024). *scdhlm: Estimating hierarchical linear models for single-case designs* (R package version 0.7.3). *Comprehensive R Archive Network* (CRAN).

Rodgers, M., Simmonds, M., Marshall, D., Hodgson, R., Stewart, L. A., Rai, D. (2021). Intensive behavioural interventions based on applied behaviour analysis for young children with autism: An international collaborative individual participant data meta-analysis. *Autism*, *25*(4), 1137–1153. https://doi.org/10.1177/1362361320985680

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. https://doi.org/10.1016/j.jsp.2013.11.005

Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Özkan, S., & Savage, M. N. (2020). Evidence-based practices for children, youth, and young adults with autism. The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, *National Clearinghouse on Autism Evidence and Practice Review Team*.

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Wilson, B. (2016). The single-case reporting guideline *in Behavioural interventions (SCRIBE) 2016 statement. Physical Therapy*, *96*(7), e1–e10. https://doi.org/10.2522/ptj.2016.96.7.e1

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). Institute of Education Sciences, U.S. Department of Education.

Wilbert, J., & Lueke, T. (2025). *scan: Single-Case Data Analyses for Single and Multiple Baseline Designs* (R package version 0.67.0). Comprehensive R Archive Network (CRAN).

Received
October 2025

Accepted
December 2025