Causal Inferences for Data in Single-Case Design Research Across Multilevel Contexts

Mikyung Shin[1], PhD, Jiyeon Park[2], PhD,

[1]Illinois State University, [2]Eastern Kentucky University

Background: A single-case design focuses on individual performance and measures the causal relationships between variables (Kazdin, 2019). This experimental design enables researchers to measure the learning behaviors of individual participants over time across phases and assess the effectiveness of an instructional strategy in improving or influencing students' progress and performance (Cooper et al., 2020). The application of multilevel modeling to single-case design research allows researchers to investigate the effects of moderators (Moayert et al., 2023), including cross-level interactions with case-level (e.g., age, gender, disability) and study-level (e.g., research design, intervention types) moderators. In many clinical settings, the outcome variables are not continuous and are often comprised of binary, categorical, and ordinal data. Additionally, researchers often encounter convergence issues during data analysis, particularly when there are excessive zeros in the dataset (Li et al., 2024), such as in the baseline phases of single-case designs (Shin et al., 2024).

Purpose/Research Question: The purpose of this paper is to provide a methodological demonstration of analyzing data in single-case design research across multilevel contexts using open-source R code and open datasets (Shin & Park, 2024). Research questions are as follows: 1. How can we measure the functional relation in multiple-probe across-subjects single-case design research? 2. How can we model a piecewise regression model to measure changes in levels and trends in single-case design? 3. How can we employ and infer causal inferences through Bayesian cumulative link mixed models?

Setting: This study was conducted in a resource room at a middle school in the southeastern region of the United States. The intervention was delivered in a one-on-one format by a special education teacher.

Participants: Four middle school students who met the following inclusion criteria participated in the study: (a) attended grades 6–8; (b) did not meet grade-level mathematics performance on the latest state-wide tests; (c) were identified as having a specific learning disability with Individualized Educational Program (IEP) goals in mathematics; and (d) demonstrated difficulty relative to the target skill, earning a score below 30% on a researcher-developed screening test.

Intervention: Technology-assisted instruction with teacher prompts was employed. This program is a web-based application comprising three main sections: Multiplication Facts, Vocabulary, and Lessons. The lessons incorporated multiple evidence-based instructional components, including explicit and systematic instruction, cognitive and metacognitive strategies, and virtual manipulatives, for teaching problem-solving strategies.

Research Design: A multiple-probe design across participants was employed to analyze the relationship between the intervention and the acquisition of word problem-solving skills by students in a time-lagged manner (Gast et al., 2018). A stability criterion was applied, where 80% of the baseline data points had to fall within ±25% of the median percentage correct for the two dependent variables before proceeding to the intervention phase (Lane & Gast, 2014).

Data Collection and Procedures: Throughout the study, a one-time screening test, a 30-minute technology training session on the basic interface of the project's web application, repeated progress monitoring assessments, fidelity of interventions, and social validity data were collected. On each intervention day, students received 30 minutes of mathematics instruction using the web-based application, followed by 10 minutes of instructional probes. After the completion of the intervention, each student took three–four maintenance tests (10 minutes per day).

Data Analysis: Two methods were employed to analyze single-case data. First, we employed visual analysis to examine data patterns within each phase of the visualization and problem-solving tasks following the guidelines of previous researchers (Cooper et al., 2020; Kratochwill et al., 2010). The proportion of non-overlapping data between adjacent conditions (baseline versus intervention and intervention versus maintenance) were calculated using Tau, a nonparametric effect size that does not adjust for baseline trends (Parker et al., 2011), where 0.20 was indicated as a small effect, 0.20 to 0.60 as a moderate effect, 0.60 to 0.80 as a large effect, and greater than 0.80 as a large to very large effect (Vannest & Ninci, 2015). We used the SingleCaseES R package (Pustejovsky et al., 2023) for this computation.

To examine the moderating effects of word problem question types, we employed Bayesian cumulative link mixed effects models and analyzed 6-point ordinal scale data for each item (Level 1) nested within repeated measurement days (Level 2) for each student (Level 3). By incorporating prior information, researchers can fit models that avoid extreme values, such as random effect variances close to zero (Meteyard & Davies, 2020) through the following Level 1 equation:

$$\text{logit}\left[\pi_{kijl}(Y > k)\right] = ln\left(\frac{\pi(Y_{ijl} > k)}{\pi(Y_{ijl} \leq k)}\right) = -\alpha_k + (\beta_{0jl} + \beta_{1jl}(t - T_{1jl}) + \beta_{2jl}(t > T_{2jl}) + \beta_{3jl}(t - T_{2jl}) \times (t > T_{2jl}) + \beta_{4jl}(t > T_{3jl}) + \beta_{5jl}(t - T_{3jl}) \times (t > T_{3jl})),$$

in which logit (π_{kijl}) is the logit link for the cumulative probability, $\pi_{kijl}(x) = P(Y > k \mid x_1, x_2, \ldots, x_p)$, of being above a particular cut point k ($k = 0, 1, \cdots 4$) for the ith fraction multiplication word problem item ($i = 1, 2, \ldots, I$) on the jth day ($j = 1, \cdots J$) for the lth student ($l = 1, 2, \cdots L$). α_k indexes thresholds, and T_{1jl}, T_{2jl}, and T_{3jl} denote three time points after which baseline, intervention, and maintenance are introduced, respectively, for the lth student. ($t > T_{2jl}$) and ($t > T_{3jl}$) are logical statements and are evaluated as 0 if false or 1 if true. Then, we added interaction terms between the intervention effects and word problem question types and considered random effects of baseline level and changes in intervention level across items, days, and students using the brms R package (Bürkner, 2017).

Findings/Results: Students demonstrated improvements from the baseline to intervention phases (Tau ranged from .76 to 1.00 for visualization and was 1.00 for problem-solving). Researchers employed Bayesian cumulative link mixed effects models to examine the moderating effects of word problem question types. Students showed greater maintenance effects on problem-solving than on visualization tasks, as reflected in changes in level (logit coefficient = 2.6) and trend (logit coefficient = 0.22).

Conclusions: Although the small number of participants limits the generalizability of the findings, single-case design research is often aimed at understanding and refining interventions on an individual level, making it particularly useful in situations where large-scale trials may not be feasible or necessary. Future research should extend this intervention and further explore its effects on students from diverse backgrounds and instructional group settings.

References

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education.

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2018). Multiple baseline and multiple probe designs. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (pp. 239–281). Routledge.

Kazdin, A. E. (2019). Single-case experimental designs. Evaluating interventions in research and clinical practice. *Behaviour Research and Therapy, 117*, 3–17. https://doi.org/10.1016/j.brat.2018.11.015

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

Li, H., Luo, W., & Baek, E. (2024). Multilevel modeling in single-case studies with zero-inflated and overdispersed count data. *Behavior Research Methods, 56*(4), 2765–2781. https://doi.org/10.3758/s13428-024-02359-7

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, Article 104092. https://doi.org/10.1016/j.jml.2020.104092

Moeyaert, M., Yang, P., & Xue, Y. (2023). Individual participant data meta-analysis including moderators: Empirical validation. *The Journal of Experimental Education, 92*(4), 723–740. https://doi.org/10.1080/00220973.2023.2208062

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*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Pustejovsky, J. E., Chen, M., & Swan, D. M. (2023). *SingleCaseES: A calculator for single-case effect sizes* (R package version 0.7.1) [Computer software].

Shin, M., Hart, S. L., & Simmons, M. (2024). Meta-analysis of single-case design research: Application of multilevel modeling. *School Psychology, 39*(6), 625–635. https://doi.org/10.1037/spq0000637

Shin, M., & Park, J. (2024). Technology-assisted instruction with teacher prompts on fraction multiplication word problems: A single-case design with visual analysis and Bayesian multilevel modeling. *Assistive Technology*. Advance online publication. https://doi.org/10.1080/10400435.2024.2415366

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development, 93*(4), 403–411. https://doi.org/10.1002/jcad.12038