MultiSCED: A tool for (meta-)analyzing single-case experimental data with multilevel modeling

Lies Declercq[1] · Wilfried Cools[1] · S. Natasha Beretvas[2] · Mariola Moeyaert[3] · John M. Ferron[4] ·
Wim Van den Noortgate[1]

Abstract

The MultiSCED web application has been developed to assist applied researchers in behavioral sciences to apply multilevel modeling to quantitatively summarize single-case experimental design (SCED) studies through a user-friendly point-and-click interface embedded within R. In this paper, we offer a brief introduction to the application, explaining how to define and estimate the relevant multilevel models and how to interpret the results numerically and graphically. The use of the application is illustrated through a re-analysis of an existing meta-analytic dataset. By guiding applied researchers through MultiSCED, we aim to make use of the multilevel modeling technique for combining SCED data across cases and across studies more comprehensible and accessible.

Keywords Single-case experimental design · SCED · Multilevel analysis · R · Shiny

In a single-case experimental design (SCED), a dependent outcome variable is manipulated and repeatedly measured within a single case, subject or unit, in order to verify the effect of the manipulations ('treatments') on the outcome (Onghena & Edgington, 2005). In the context of SCED data, there is a variety of different ways to quantify the effectiveness of a treatment. Depending on the specific SCED data characteristics and the research question under investigation, some techniques are more appropriate than others. A strong tradition of visual analysis exists and still predominates in the field (Manolov & Moeyaert, 2017). Over the years, researchers' questioning of the reliability and consistency of visual analysis (Parsonson & Baer, 1992) and their advocating for complementing visual analysis with statistical analysis techniques (Fisch, 2001; Manolov

& Solanas, 2013) have led to methodological work on developing effect sizes quantifying the effectiveness of an intervention.

One of the proposed methods that has gained more attention throughout the years is multilevel modeling (Jamshidi et al. 2018). Multilevel modeling (also referred to as hierarchical linear or linear mixed modeling) is a very flexible approach for combining and summarizing SCED data across cases and across studies, taking into account the hierarchical nature of SCED data. The application of multilevel modeling for meta-analyzing SCED data was first proposed by Van den Noortgate and Onghena (2003a) and since then this approach has been explored and extended in multiple methodological works, e.g., Ferron, Farmer, & Owens, (2010), Owens and Ferron (2012), Moeyaert et al. (2017), Baek and Ferron (2013), Rindskopf and Ferron (2014), Joo et al. (2017) and Ugille et al. (2012). However, despite these efforts, it has been shown that the approach is only slowly gaining attention among SCED data analysts and meta-analysts. There exists a gap between statistical advances and the actual practices applied by behavioral researchers (Manolov & Moeyaert, 2017). This seems to be especially true for multilevel modeling (Jamshidi et al. 2018), which requires a good understanding of regression model building, the associated parametric assumptions, and the interpretation of the resulting estimates.

To bridge this gap between methodological and applied research, free and easy-to-use software tools and tutorials

✉ Lies Declercq
lies.declercq@kuleuven.be

[1] Faculty of Psychology and Educational Sciences,
imec - ITEC, KU Leuven, Leuven, Belgium

[2] Department of Educational Psychology, University of Texas,
Austin, TX 78712, USA

[3] Department of Educational Psychology and Methodology,
University at Albany-State University of New York, Albany,
NY, USA

[4] Department of Educational Measurement and Research,
University of South Florida, Tampa, FL 33620, USA

are crucial (Manolov & Moeyaert, 2017; Shadish, 2014). Specifically for multilevel modeling, applied researchers can already rely on dedicated software such as MLwiN (Rasbash et al., 2009) and HLM (Raudenbush, Bryk, & Congdon, 2013). Common general statistical software packages also offer multilevel modeling functionality: SPSS provides the MIXED command (Peugh & Enders, 2005), SAS has the MIXED procedure (Littell et al., 2007) and for R there are packages such as lme4 (Bates et al., 2015) or nlme (Pinheiro et al., 2018). However, even though the existing software for multilevel modeling is suitable to respond to the needs of applied SCED researchers searching for data analysis tools, applying multilevel modeling to summarize SCED data may often not be straightforward given the specific SCED data characteristics. Correctly implementing a multilevel model, obtaining parameter estimates, testing hypotheses, interpreting the results and creating meaningful graphs of them can be quite complex.

MultiSCED (available at http://www.single-case.com/MultiSCED) is a web application developed to address these issues by guiding researchers through the steps of performing a multilevel analysis of their SCED data. MultiSCED is built with Shiny (Chang et al., 2017), a framework to create interactive web applications that provide an interface to R functionality (R Core Team, 2013). The application offers a point-and-click user interface, allowing practitioners to use the freely available R software environment without having to know or hard code the R syntax. The goal of MultiSCED is to help users understand how to use multilevel modeling for their SCED data analysis. The application tries to do this in three ways. First, MultiSCED is designed to perform data analysis step by step by starting with a simple and familiar (single level) linear regression model. After exploring the data and the single-level regression results case by case, multiple cases' data can be combined using a multilevel model. Second, MultiSCED provides the results not only numerically but also provides a graphical presentation, and this very much facilitates interpretation. The obtained graphs can be saved as figures to use in data analysis reports. Finally, MultiSCED automatically displays the multilevel modeling equations, providing the user immediately with the correct mathematical representation of the model he/she has built. The underlying R code formula required to fit the model is also displayed. This code provides a starting point to users who are interested in extending the analyses within R.

The aim of this paper is to introduce MultiSCED and to give a brief overview of the options embedded in the application. For a more detailed description on how to navigate through the point-and-click user interface, we refer to the MultiSCED user guide, which is available at https://kuleuven.box.com/v/MultiSCEDUserGuide. This user guide illustrates all features of MultiSCED through a real data example based on a dataset by Shogren, Faggella-Luby, Bae and Wehmeyer (2004). This example is also included in this paper in a more summarized version, along with screenshots of the results as displayed within the MultiSCED environment. The illustration is at an introductory level aimed at applied SCED data analysts and meta-analysts without prior experience with multilevel modeling. Starting with a single-level analysis per case, we work our way up to a two-level model combining SCED data across cases within one study and finally to a full three-level model combining SCED data across cases and across studies using one statistical analysis technique. In conclusion, we highlight the application's strengths and argue how MultiSCED can help to bridge the aforementioned gap between methodological and applied SCED research.

Empirical demonstration of MultiSCED

Illustrative dataset

For illustrating the use of MultiSCED, we chose a meta-analytic dataset collected by Shogren et al. (2004). In their paper, the authors collect and summarize 13 single-case studies with a total of 30 participants in a meta-analysis. They were interested in summarizing empirical evidence to evaluate the effectiveness of choice-making interventions on problem behavior for people with disabilities. In the original study, the authors used two non-overlap indices to analyze the effect of the treatment. Van den Noortgate and Onghena (2008) retrieved the raw data from graphs in the primary studies identified by Shogren et al. (2004) and re-analyzed this dataset using multilevel modeling. The same dataset is used in this paper and is available at https://kuleuven.app.box.com/v/Shogren2004.

Data preparation

MultiSCED has several prerequisites that should be met before uploading a dataset into the application. In this paragraph, we list these prerequisites and verify them for the Shogren et al. (2004) illustrative dataset. First, the data should be collected by means of AB-phase designs, with multiple measurements per case during a baseline and a treatment phase. Several primary studies in the Shogren et al. (2004) meta-analysis used reversal designs. For these studies, only data from the first baseline and immediately following intervention phase are included. Although more complex SCED models are beyond the scope of Multi-

SCED, we want to encourage more advanced users to implement multilevel models for extended designs (e.g., reversal or alternating treatment designs) directly in R with lme4, using for example the work by Moeyaert et al. (2015), Shadish, Kyse, and Rindskopf (2013) or Van den Noortgate and Onghena (2003b) as a guideline.

For analysis in MultiSCED, the data should include at least the following variables: a variable indicating the measurement occasion (e.g., time, day or session), a variable indicating the phase (whether the measurement occasion is part of the baseline or the treatment phase), a variable indicating the case, subject, or participant, and finally the outcome variable (the value of the dependent variable of interest for a particular measurement occasion). An excerpt of the data file is shown in Fig. 1. The data file is stored in .txt format, the only supported format for upload in MulitSCED. Each row represents an observation for a case and each column represents a variable.

The outcome of interest (i.e., the dependent variable, denoted as Y in Fig. 1) is problem behavior. To avoid dependency due to multiple outcomes per participant, only one outcome indicating (overall) problem behavior was selected when multiple outcomes per participant (e.g., problem behavior and aggressive behavior) were reported. Multilevel analysis of SCED data with multiple outcomes is beyond the scope of MultiSCED because this involves either a multivariate multilevel model (rather than the univariate models illustrated in MultiSCED and in this paper) or a nested or cross-classified model with a fourth level for the outcome. For users interested in taking the dependency between multiple outcomes into account, we refer to Van

den Noortgate et al. (2014) as a starting point. Furthermore, data for one participant from the study by Romaniuk et al. (2002) was excluded because for this participant, two different conditions were alternated within the reversal design. Finally, data for one participant from the study by Kern et al. (2001) were also excluded because for this participant the only measured outcome was one of positive behavior. Some of the outcomes measured in the primary studies of the Shogren et al. (2004) meta-analysis were not measured on a continuous scale (e.g., in the study by Dibley and Lim, 1999), violating the normality assumption underlying the linear multilevel models provided in MultiSCED. An alternative to account for discrete data is to use a generalized linear mixed model instead, for which we refer to Declercq et al. (2018).

Additional variables in the dataset and their coding protocol are listed in Table 1. In the original meta-analysis by Shogren et al. (2004), more independent variables are included, but these are omitted here, as an extensive moderator analysis is beyond the scope of this paper.

One-level analysis

SCEDs have been introduced as a way to closely study whether individual cases respond to a treatment or not (Shadish, Rindskopf, & Hedges, 2008). An ordinary least squares (OLS) regression analysis on a particular case's data is a way to quantify the effect of the treatment for the individual of interest. In MultiSCED, this is referred to as a 'one-level analysis' and it yields case-specific treatment effect estimates. Including Phase and Time and

Author	Name	Phase	Time	Y	Gender	Age
Bambara1995	Al	0	1	3	0	50
Bambara1995	Al	0	2	3	0	50
Bambara1995	Al	0	3	2	0	50
Bambara1995	Al	0	4	4	0	50
Bambara1995	Al	1	5	1	0	50
Bambara1995	Al	1	6	1	0	50
Bambara1995	Al	1	7	0	0	50
Bambara1995	Al	1	8	0	0	50
Cole2002	Keith	0	−6	17	0	8
Cole2002	Keith	0	−5	8	0	8
Cole2002	Keith	0	−4	14	0	8
Cole2002	Keith	0	−3	62	0	8
Cole2002	Keith	0	−2	23	0	8
Cole2002	Keith	0	−1	39	0	8
Cole2002	Keith	1	0	31	0	8
Cole2002	Keith	1	1	12	0	8
Cole2002	Keith	1	2	0	0	8
Cole2002	Keith	1	3	0	0	8
Cole2002	Keith	1	4	20	0	8
Cole2002	Keith	1	5	15	0	8
Cole2002	Keith	1	6	0	0	8
Cole2002	Keith	1	7	29	0	8
Cole2002	Keith	1	8	15	0	8

Fig. 1 Shogren et al. (2004) input data file. The meta-analytic data are stored in a tab-delimited text (.txt) file with each row representing an observation and each column representing a variable

Table 1 Variables in the illustrative dataset based on the meta-analysis by Shogren et al. (2004)

Variable	Type	Description
Author	Text	First author and year
Name	Text	Participant name
Time	Numeric	Measurement occasion
Phase	Factor	0 - baseline
		1 - treatment
Y	Numeric	Problem behavior outcome
Gender	Factor	0 - male
		1 - male
Age	Numeric	Participant age

their interaction as independent variables, we define a one-level model for each individual case in the Shogren et al. (2004) dataset as

$$Y_i = \beta_0 + \beta_1 \text{Time}_i + \beta_2 \text{Phase}_i + \beta_3 (\text{Time}_i \times \text{Phase}_i) + e_i \quad (1)$$

The subscript i denotes the measurement nested within the case. The variable Time_i denotes the time at which measurement i was observed. This can be real time (e.g., days or minutes) or a time indication (e.g., session number). The residuals e_i are assumed to be independent, following a normal distribution with zero mean and standard deviation σ_e. In this one-level model, the only source of variation is the sampling errors e_i: this is the random variation of the sample measurements around the expected value. When a measurement Y_i is part of the baseline phase, the dummy variable score Phase_i equals 0. Equation 1 then simplifies to

$$Y_i = \beta_0 + \beta_1 \text{Time}_i + e_i \quad (2)$$

for the baseline phase data's model. Thus the baseline model is a straight line with respect to Time, with an intercept β_0 and a slope β_1. For measurements Y_i in the treatment phase, we have that Phase_i equals 1 and then Equation 1 can be written as

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 \text{Time}_i + \beta_2 + \beta_3 \text{Time}_i + e_i \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \text{Time}_i + e_i \end{aligned} \quad (3)$$

The intercept in the treatment phase (i.e., the predicted outcome score at the start of the treatment phase) is $\beta_0 + \beta_2$ and the slope for the treatment phase is $\beta_1 + \beta_3$. Thus β_2 is the effect of the treatment on the intercept, and β_3 is the effect of the treatment on the slope. To facilitate the interpretation of β_0 and β_2 in particular, MultiSCED offers an option to center the time variable around the time at the first measurement occasion in the treatment phase (Fig. 2). When the data are centered, β_0 refers to the score at the first measurement occasion after the baseline phase as

predicted if the baseline trend is projected into the treatment phase, whereas $\beta_0 + \beta_2$ refers to the score at the same measurement occasion but predicted by the treatment phase data's trajectory. In other words, β_2 refers to the immediate effect of the intervention when the treatment phase starts.

An important issue when combining data across studies is that different studies may use different metrics to measure the outcomes. This might result in different measuring scales used for the dependent variable. To make outcomes from different studies comparable, it is required to standardize the outcome variable before combining the outcomes in a multilevel analysis. MultiSCED offers the option to standardize the data according to the method proposed by Van den Noortgate and Onghena (2008). In this method, an OLS regression is performed on each individual case, according to the model defined above (1). Next, the individual scores Y_i are standardized by dividing them by the estimated residual within-subject standard deviation $\hat{\sigma}_e$ of case j from study k:

$$Y_{i(jk)}^{\text{ST}} = \frac{Y_{i(jk)}}{\hat{\sigma}_{e(jk)}} \quad (4)$$

This procedure is applied to the choice-making intervention dataset and all results discussed in the remainder of this paper have been obtained based on standardized problem behavior scores.

The results of the one-level OLS regression are shown as a table in Fig. 3 and are displayed graphically in Fig. 4 using screenshots of the MultiSCED environment. Each row of the output table represents a case-specific regression coefficient estimate together with its standard error, t value and p value. For instance for Brooke (Fig. 3), a case from the study by Romaniuk et al. (2002), we see that the estimated intercept ($\hat{\beta}_0 = 5.09$) and the Time and Phase coefficients ($\hat{\beta}_1 = 0.37$ and $\hat{\beta}_2 = -3.93$) are statistically significant ($p < 0.01$ for the intercept and Phase, and $p < 0.05$ for Time, based on two-tailed significance testing), while the estimated interaction of Time and Phase ($\hat{\beta}_3 = -0.60$) is not statistically significant ($p = .11$ based on two-tailed significance testing). The one-level regression results imply that if the baseline continued, Brooke would have had a (standardized) score of 5.09 on problem behavior at the start of the treatment phase, and this score increases by 0.37 with each additional time unit. The choice-making intervention has a statistically significant immediate effect on Brooke's problem behavior, which is decreased by 3.93 on the standardized problem behavior scale. The time trend in the treatment phase is estimated to be $0.37 - 0.60 = -0.23$. However, because of the non-significant p value associated with the interaction between Time and Phase, we cannot conclude that the treatment effect on the time trend is statistically significantly different from zero.
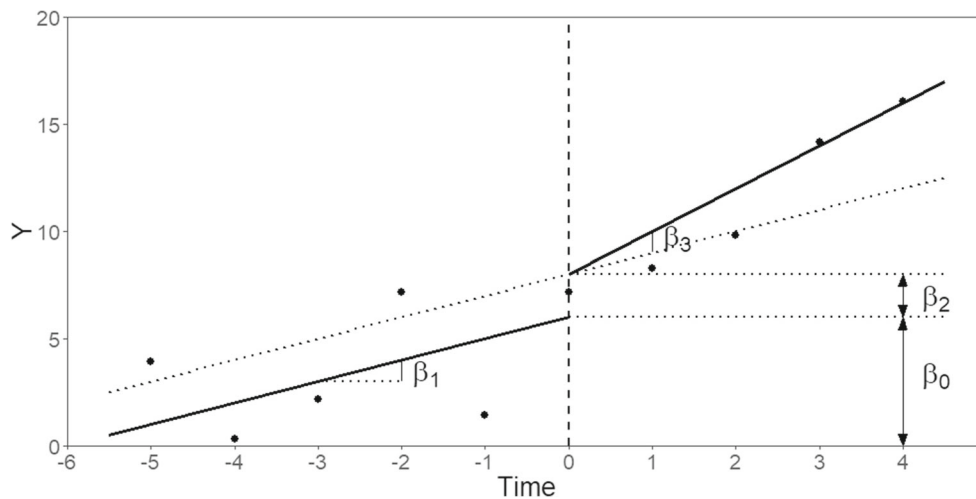
Fig. 2 Graphical interpretation of the one-level model parameters from Eq. 1 after centering the time variable. The immediate treatment effect is expressed by β_2 and the effect of the treatment on the time trend is expressed by β_3

We repeated the one-level analysis with unstandardized data to be able to interpret the fixed effects estimations within the context of the specific study, i.e., on the original scale. For Brooke, the estimated unstandardized intercept was $\hat{\beta}_0 = 89.67\%$ [$t = 7.001$, $p < 0.001$], indicating that at the time of the beginning of the intervention, Brooke showed problem behavior for 89.67% of the 5-min total session time. Without the treatment, Brooke's percentage of session time in which she displayed problem behavior increased by $\hat{\beta}_1 = 6.53\%$ [$t = 2.87$, $p = 0.02$] every session. The immediate effect of the treatment was a $\hat{\beta}_2 = 69.27\%$ [$t = -3.70$, $p < 0.01$] drop in problem behavior within sessions and this continued to decrease by $\hat{\beta}_3 = 10.63\%$ [$t = -1.77$, $p = 0.11$] every session.

The regression lines based on the standardized data for Brooke are shown as a MultiSCED screenshot in Fig. 4. Recall that we chose to center the data (i.e., transform the time variable so that Time = 0 for the first observation in the treatment phase). This was done deliberately so that the β_2 coefficient of the `Phase` variable from the model in Eq. 1 represents the jump between the intercept based on the baseline data's trajectory and the intercept based on the treatment data's trajectory in Fig. 2. For Brooke, a substantial drop in problem behavior at the start of the treatment phase is clearly visible in Fig. 4, and this drop is

statistically significant according to the p value in Fig. 3. The upward slope in the baseline phase (0.37) changes to the slightly downward slope (-0.23) in the treatment phase. This effect was non-significant, meaning that true slope in Brooke's treatment phase could actually still be identical to the upward slope in her baseline phase.

Also shown in Fig. 4 are a 95% confidence band (in dark grey) and a 95% prediction band (in light grey). Both bands are centered around the regression line. At a given time point, we have a 95% probability that the regression line will pass between the two values of the confidence interval, whereas we have a 95% probability that a hypothetical, non-observed score would be situated between the two values of the prediction interval. The confidence band is less wide, because it only addresses uncertainty about the expected mean score given the time and phase, whereas the prediction band addresses uncertainty about individual observations.

Effect size predictions

Based on the estimates of the one-level regression coefficients, we can estimate the effect for a given time point. Assuming the relation as estimated would be persistent in time, we can even make predictions for future time points beyond the time of the last measurement. For Brooke, the

Name	Regressor	Coefficient	Standard Error	t-value	p-value
Brooke	(Intercept)	5.088	0.727	6.996	0.000
Brooke	Time	0.371	0.129	2.870	0.017
Brooke	Phase1	-3.934	1.063	-3.701	0.004
Brooke	Time:Phase1	-0.603	0.342	-1.763	0.108

Fig. 3 One-level regression results table and effect size prediction for Brooke from the study by Romaniuk et al. (2002). The one-level model was estimated based on standardized scores. Note that the p value for the intercept indicates that $p < 0.01$ (rather than $p = 0$)
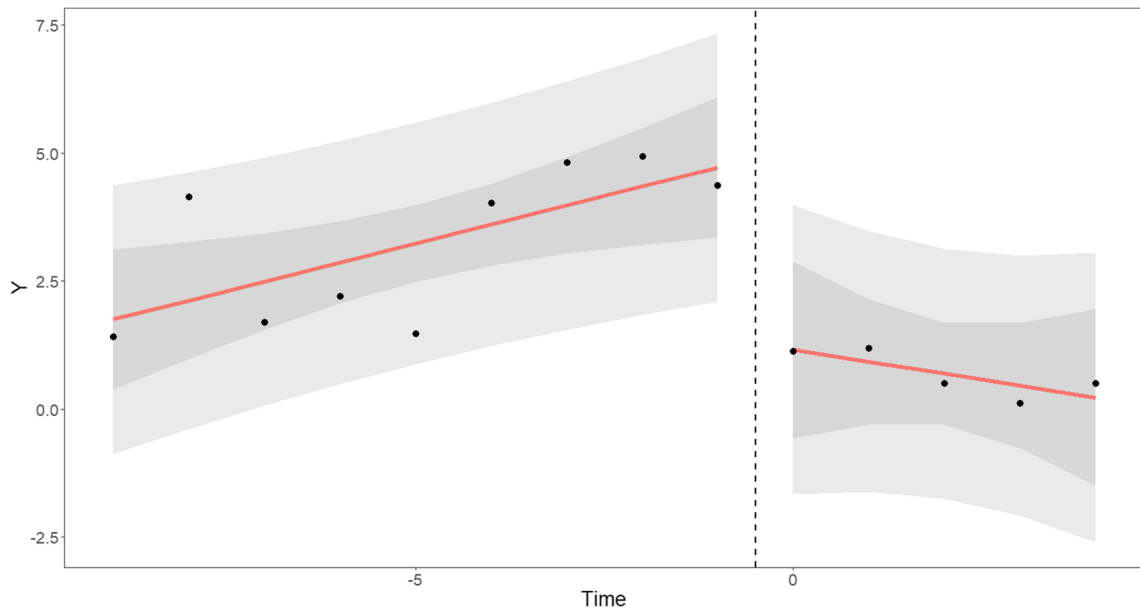
Fig. 4 One-level regression results plot for Brooke from the study by Romaniuk et al. (2002). The one-level model was estimated based on standardized scores. Prediction bands (*light grey*) and confidence bands (*dark grey*) were calculated for $\alpha = .05$

last measurement is at Time $= 5$ (remember that earlier we chose to center all observations so that Time $= 0$ occurs at the first measurement in the treatment phase). The estimated standardized score at that time is $5.09 + 0.37 \times 5 - 3.93 - 0.60 \times 5 = 0.01$. If the intervention had not taken place, the standardized score is estimated to be $5.09 + 0.37 \times 5 = 6.94$. By taking the difference between these two, we can get an idea of the estimated overall effect at Time $= 5$ or, by extension, at any time point beyond that. For Brooke,

this effect size at Time $= 5$ equals $0.01 - 6.94 = -6.93$. Note that this is negative because there is a drop in problem behavior due to the intervention (as expected). Going back to the regression equations (Eqs. 2 and 3) and taking the difference between them, we get the following equation expressing the size of the effect at any point in time:
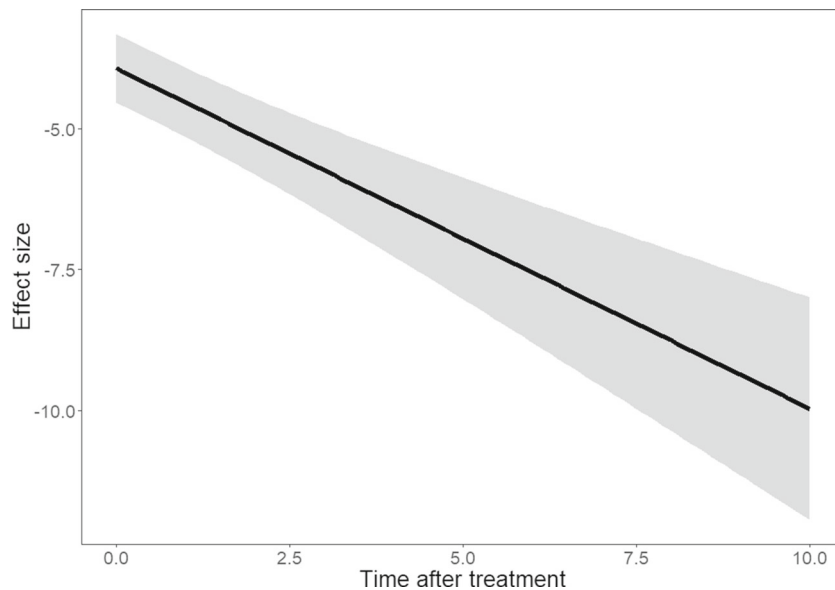
$$\beta_2 + \beta_3 \text{Time}_i \tag{5}$$

Fig. 5 Effect size prediction plot based on the one-level regression analysis for Brooke from the study by Romaniuk et al. (2002). The trajectory is given by Eq. 5. The associated confidence band was calculated for $\alpha = .05$

Note that for the Shogren et al. (2004) dataset, a time trend model is used rather than the intercept-only model (without a slope) used by Van den Noortgate and Onghena (2008). In their paper, Van den Noortgate and Onghena (2008) illustrate how the coefficient for the treatment effect on the intercept can be interpreted as the standardized difference between the average response in the baseline and in the treatment phase (i.e., a standardized mean difference). For the time trend model illustrated in this paper, the treatment effect is expressed by two regression coefficients instead of one, and the mean difference between the baseline and treatment phase depends on the time point (see Eq. 5). The fixed effect expressing the treatment effect on the intercept alone can therefore not be interpreted as a standardized mean difference.

In MultiSCED, the effect size predictions and their corresponding confidence interval can be shown in table format for a specific time point after the treatment, or in a plot for a series of time points after the treatment (Fig. 5). Note in Fig. 5 that for predictions further in the future the confidence band is getting wider, reflecting the increasing uncertainty about the size of the effect.

Two-level analysis

In addition to the case-specific results, we may want to synthesize the data at hand to investigate whether the treatment effects can be generalized across multiple cases within a study (Moeyaert et al., 2014). Building further on Eq. 1, we add an index j to denote case j within the study:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{Time}_{ij} + \beta_{2j}\text{Phase}_{ij}$$
$$+ \beta_{3j}\left(\text{Time}_{ij} \times \text{Phase}_{ij}\right) + e_{ij} \qquad (6)$$

The standard deviation σ_e of the residuals e_{ij} (again assumed to be independent and normally distributed) is assumed to be identical for all cases. Each of the regression coefficients in Eq. 6 is split up into a fixed effect γ plus a random case-specific deviation u, called a random effect:

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} + u_{1j} \\ \beta_{2j} = \gamma_{20} + u_{2j} \\ \beta_{3j} = \gamma_{30} + u_{3j} \end{cases} \qquad (7)$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & & & \\ \sigma_{u_0 u_1} & \sigma_{u_1}^2 & & \\ \sigma_{u_0 u_2} & \sigma_{u_1 u_2} & \sigma_{u_2}^2 & \\ \sigma_{u_0 u_3} & \sigma_{u_1 u_3} & \sigma_{u_2 u_3} & \sigma_{u_3}^2 \end{pmatrix} \right]$$

The random effects are assumed to be multivariate normally distributed. In the two-level model (6 and 7), researchers are often interested in γ_{20} and γ_{30}, which express the average effect of the treatment on respectively the intercept and the slope of the regression line in the baseline phase.

Furthermore, they are interested in the variance components $\sigma_{u_2}^2$ and $\sigma_{u_3}^2$, because these express the extent to which these effects vary between cases within the study. Note that although in this case we include random effects for all regression parameters, MultiSCED allows to choose for which specific parameters to include a random effect and as such the covariance structure can be manipulated.

For the study by Romaniuk et al. (2002) the numerical results of the two-level analysis are shown in Fig. 6. Across cases, the expected standardized score increases on average by 0.44 per time unit [$t(62.99) = 8.11$, $p < 0.001$] if there is no intervention, up until 6.48 points [$t(5.40) = 15.41$, $p < .001$] at the start of the treatment phase (Fig. 6). The intervention has an average immediate effect of -2.63 points [$t(4.07) = -2.55$, $p = 0.06$] and it decreases the time trend by -0.50 points [$t(31.52) = -3.80$, $p < 0.001$]. Standard errors for the intercept and the time trend during the baseline are relatively small, resulting in quite high t values (the ratios of the estimates over the corresponding standard errors). The baseline intercept, baseline trend, and effect of the treatment on the trend are significant at a $\alpha = 5\%$ level, indicating that the treatment might not have had a significant immediate effect (although it is significant at the $\alpha = 10\%$ level), but that it did have an impact on the slope in the treatment phase.

Hypothesis testing for the fixed effects of multilevel models in MultiSCED is based on a t test with the Kenward–Roger approximation for the degrees of freedom (Kenward & Roger, 1997). Although this is the recommended method for Wald-type tests on multilevel modeling (Ferron, Farmer, & Owens, 2010; Luke, 2017), the assumptions related to the asymptotic distribution of the test statistic in a multilevel model are often not met (Bates, 2006; Bolker et al., 2018; Luke, 2017). Alternative testing approaches (currently not implemented in MultiSCED) include likelihood ratio tests (Welham & Thompson, 1997), likelihood profile tests, Markov-chain Monte Carlo (MCMC) sampling (Baayen, Davidson, & Bates, 2008) or bootstrapping methods.

Again, we repeat the analysis for the unstandardized Romaniuk et al. (2002) data to facilitate interpretation within the context of the study. Overall, if the seven study participants' baselines had been extended to the time of the first treatment phase observation, they would have on average been expected to show problem behavior during $\hat{\beta}_0 = 88.80\%$ [$t(11.89) = 18.70$, $p < 0.001$] of the 5-min total session time on average at the end of the baseline phase. Without the treatment, the percentage of time the participants showed problem behavior increased by $\hat{\beta}_1 = 6.10\%$ across sessions [$t(8.27) = 6.92$, $p < 0.001$]. After the introduction of choice-making interventions, the percentage of problem behavior dropped to $\hat{\beta}_0 + \hat{\beta}_2 = 88.80\% - 39.71\% = 49.09\%$ [$t(4.00) = -2.54$, $p = .06$ for $\hat{\beta}_2$] and continued to decrease by $\hat{\beta}_3 = 6.94\%$ [$t(4.03) = -3.06$, $p = 0.04$].

Fixed effects

```
                Estimate Std. Error      KR df    t value       p value
(Intercept)   6.4783938 0.42053733   5.404880  15.405038  1.131973e-05
Time          0.4435164 0.05466193  62.989433   8.113808  2.259034e-11
Phase1       -2.6260150 1.03170047   4.070857  -2.545327  6.255667e-02
Time:Phase1  -0.5026759 0.13232600  31.521110  -3.798769  6.248829e-04
```

Random effects

```
Groups    Name         Std.Dev.    Corr
Name      (Intercept)  0.63658283
          Time         0.00025594  -1.000
          Phase1       2.08552985   1.000 -1.000
          Time:Phase1  0.04225832   1.000 -1.000   1.000
Residual               0.96023119
```

Fig. 6 Two-level regression results tables for the study by (Romaniuk et al. 2002). The two-level model was estimated based on standardized scores

Figure 7 gives a graphical presentation of the two-level regression results for each of the cases of the study by Romaniuk et al. (2002): the green line represents the average study trajectory and is identical across cases, while the yellow lines represent each individual participant's trajectory. Using prior empirical information from the other participants, these individual participant's trajectories are drawn based on the empirical Bayes' estimates of the participant-specific coefficients. In the baseline phase, there is not much variation across cases around the study mean: all yellow lines are very close to the green line. The slopes are nearly identical and the intercepts vary only slightly. This reflects the small variance estimates for these two random effects in Fig. 7, compared to the size of the corresponding fixed effects: $\sigma_{u_0}^2 = 0.64^2$ for the intercept (where $\gamma_{00} = 6.48$) and $\sigma_{u_1}^2 = 0.0003^2$ for Time (where $\gamma_{10} = 0.44$). The two-level analysis trajectories for each individual case, as estimated by the empirical Bayes method, are closer to the observed values, while the study's overall mean trajectory incorporates information from all cases and is therefore sometimes further away from the observed values. This difference is clearly illustrated in Brooke's treatment phase. In case the empirical Bayes trajectories (i.e., the dark yellow lines on Fig. 7) are far from the observed data, this might indicate a model misspecification. In Baek et al. (2016) for example, a similar two-level model including a trend is fit. Upon inspection of the individual Bayes trajectories (p. 21, Figure 2), the authors state that 'because, overall, the visual result of the first model appeared to be poor,' they

are motivated to 'consider an alternative model that may represent the data better.' (p. 22)

From Fig. 7 we can deduce that there is slightly more variation in regression lines across cases in the treatment phase than there is in the baseline phase. The immediate decrease in problem behavior is quite large for Brooke, Gary, and Maggie, but not so much for Christy and Rick. The average immediate drop in problem behavior of −2.63 points has a relatively large associated standard deviation of 2.09 (Fig. 6) for the Phase random effects, which reflects this between-case variation. The slopes in the treatment phase are more similar and show a slightly decreasing trend, which does not vary much across cases. The corresponding random effects' standard deviation is 0.04. Finally the residual or within-case standard deviation of 0.96 approximates 1 because we standardized the data by dividing each participant's scores by the root mean squared error.

The correlations between the random effects are typically hard to estimate for single-case design data. The estimates all equal 1 or −1 in our example (see Fig. 6). Results like these often reflect that there might have been a model misspecification or that there simply might have been too little information to estimate the parameters. Still, the signs of the correlations correspond to what we see in the graphs (see Fig. 7). For example, the positive correlation between the intercept and Phase residuals indicates that when a case has a baseline intercept that is large (resp. small) compared to the study's average baseline intercept, its treatment intercept will generally be also large (resp. small) compared
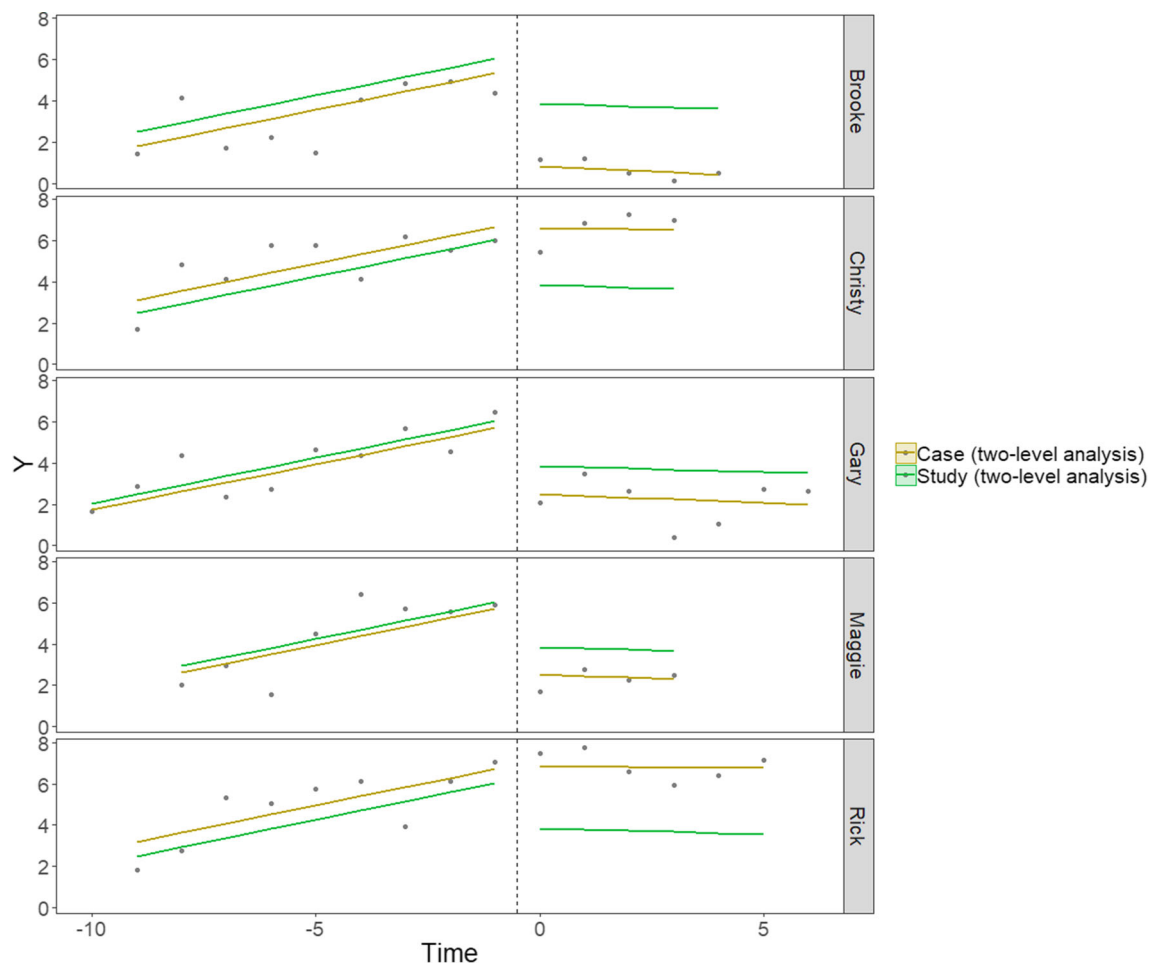
Fig. 7 Two-level regression results plot for the study by Romaniuk et al. (2002). The two-level model was estimated based on standardized scores

to the study's average treatment intercept. This is also reflected in Fig. 7: participants whose case-specific baseline trajectory (i.e., the yellow line) is above (resp. below) the study baseline trajectory (i.e., the green line) all have a case-specific treatment trajectory that is also above (resp. below) the study treatment trajectory.

Comparison to one-level analysis results

The case-specific means of the two-level analysis can be compared with the means obtained for each case in the one-level analysis. Recall from the one-level model in Eq. 1 that the regression line obtained through the one-level analysis is given by

$$\hat{\beta}_0 + \hat{\beta}_1 \text{Time}_i + \hat{\beta}_2 \text{Phase}_i + \hat{\beta}_3 \left(\text{Phase}_i \times \text{Time}_i \right) \quad (8)$$

for case j. For the cases from the study by Romaniuk et al. (2002), these regression lines are plotted in Fig. 8 in

red for the one-level case-specific mean and in yellow for the two-level case-specific mean. In the two-level model, the case-specific regression lines are considered as varying randomly around the study regression line, plotted as the green line in Fig. 8. This line's equation is given by

$$\hat{\gamma}_{00} + \hat{\gamma}_{10} \text{Time}_i + \hat{\gamma}_{20} \text{Phase}_i + \hat{\gamma}_{30} \left(\text{Phase}_i \times \text{Time}_i \right). \quad (9)$$

The one-level model estimates a particular case's regression line (in red) by only using data from that particular case, which is why the red lines follow the observed scores more closely. The two-level model however combines information from all cases within the study. The particular case's regression line is considered as varying randomly around a study mean. Hence the yellow lines deviate slightly more from the underlying case's observed scores, because they are drawn towards the green line, i.e., the study mean. This is precisely how empirical Bayes estimation of the participant-specific coefficients works: the two-level case regression lines (in green) are a weighted combination
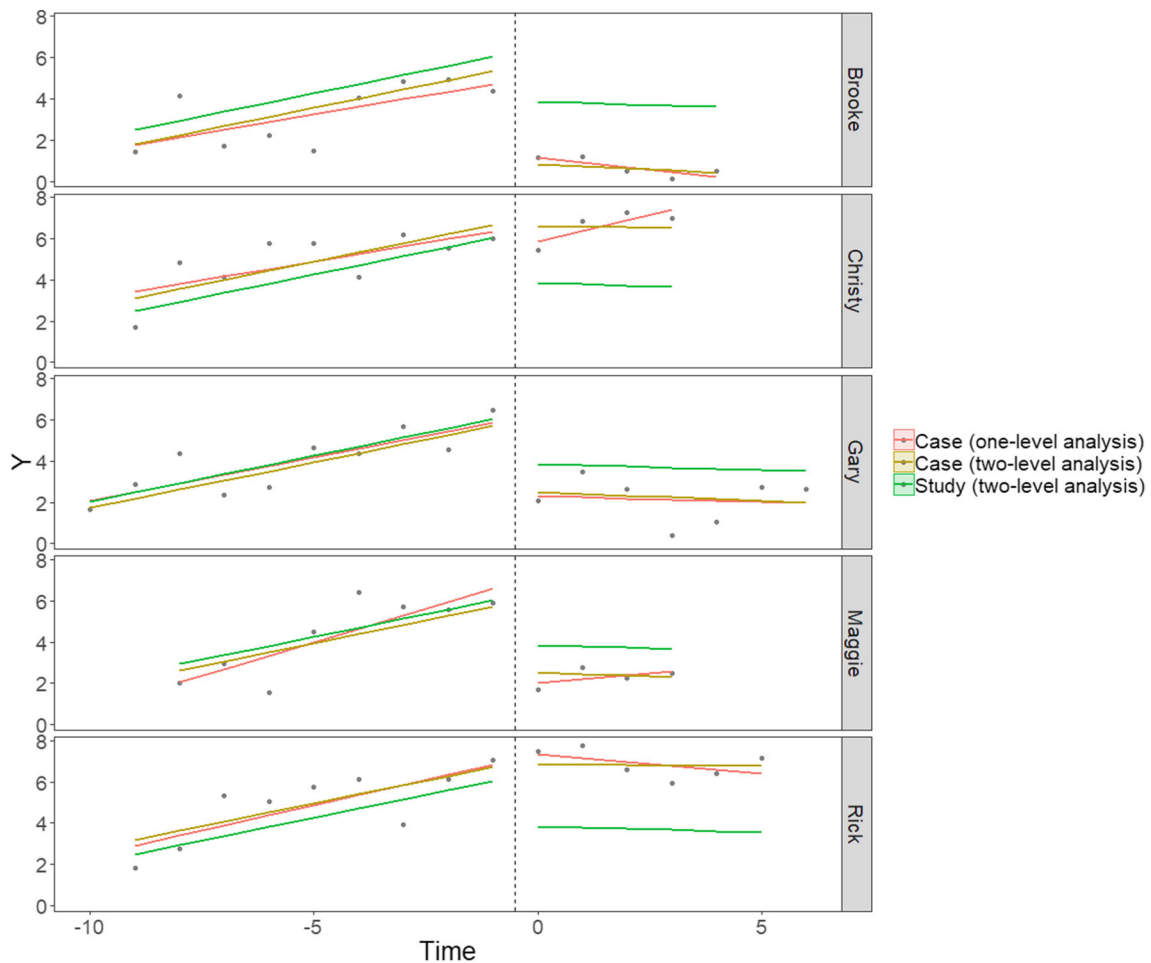
Fig. 8 Comparison of one- and two-level regression analyses for the study by Romaniuk et al. (2002). Both models were estimated based on standardized scores. MultiSCED has options to plot any combination of the displayed mean trajectories

of the one-level OLS regression lines (in red) and the overall study mean regression line (in green).

Three-level analysis

To conduct a meta-analysis of the Shogren et al. (2004) dataset, we now combine all 30 cases across all 13 studies in order to further generalize the treatment effect estimates. The two-level model is extended to a three-level model, with measurements at the first level, cases at the second level, and studies at the third level. The regression equation for the two-level model (Eq. 6) is extended by adding an index k to denote the study:

$$
\begin{aligned}
Y_{ijk} = {} & \beta_{0jk} + \beta_{1jk}\mathrm{Time}_{ijk} + \beta_{2jk}\mathrm{Phase}_{ijk} \\
& + \beta_{3jk}\left(\mathrm{Time}_{ijk} \times \mathrm{Phase}_{ijk}\right) + e_{ijk}
\end{aligned} \tag{10}
$$

Again the residuals e_{ijk} are assumed to be independent and normally distributed with the standard deviation σ_e, which is assumed to be identical for all cases. At the second level, the coefficients from the first level vary across cases around a study-specific mean $\theta_{.k}$ by a random case effect $u_{.jk}$:

$$
\begin{cases}
\beta_{0jk} = \theta_{00k} + u_{0jk} \\
\beta_{1jk} = \theta_{10k} + u_{1jk} \\
\beta_{2jk} = \theta_{20k} + u_{2jk} \\
\beta_{3jk} = \theta_{30k} + u_{3jk}
\end{cases}
$$

$$
\begin{pmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & & & \\ \sigma_{u_0 u_1} & \sigma_{u_1}^2 & & \\ \sigma_{u_0 u_2} & \sigma_{u_1 u_2} & \sigma_{u_2}^2 & \\ \sigma_{u_0 u_3} & \sigma_{u_1 u_3} & \sigma_{u_2 u_3} & \sigma_{u_3}^2 \end{pmatrix} \right] \tag{11}
$$

At the third level, the coefficients from the second level vary across studies around an overall mean $\gamma_{...}$ (i.e., the fixed effect) by a random study effect $v_{..k}$:

$$
\begin{cases}
\theta_{00k} = \gamma_{000} + v_{00k} \\
\theta_{10k} = \gamma_{100} + v_{10k} \\
\theta_{20k} = \gamma_{200} + v_{20k} \\
\theta_{30k} = \gamma_{300} + v_{30k}
\end{cases}
$$

$$
\begin{pmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v_{30k} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{v_0}^2 & & & \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 & & \\ \sigma_{v_0 v_2} & \sigma_{v_1 v_2} & \sigma_{v_2}^2 & \\ \sigma_{v_0 v_3} & \sigma_{v_1 v_3} & \sigma_{v_2 v_3} & \sigma_{v_3}^2 \end{pmatrix} \right] \quad (12)
$$

Again, we are mainly interested in the overall averages for the treatment effect on the intercept and the slope, as expressed by the fixed effects γ_{200} and γ_{300}, in their between-case variances $\sigma_{u_2}^2$ and $\sigma_{u_3}^2$, and in their between-study variances $\sigma_{v_2}^2$ and $\sigma_{v_3}^2$.

The output table for the three-level analysis of the Shogren et al. (2004) data as obtained with MultiSCED is presented in Fig. 9. The results are presented graphically in Fig. 10. From Fig. 9, we see that the overall average baseline intercept is 4.74 [$t (12.44) = 5.75$, $p < 0.001$] and the average baseline slope is 0.21 [$t (10.59) = 3.36$, $p = 0.007$]. The overall immediate treatment effect is

-3.60 [$t (14.23) = -4.16$, $p < 0.001$] and the overall effect on the slope is -0.25 [$t (8.20) = -3.97$, $p = 0.004$]. Based on the evidence from the 13 studies, it seems that providing choice-making opportunities might reduce problem behavior in people with disabilities as much as four times $(4.74/(4.74 - 3.60) = 4.16)$ in a very short amount of time. Overall, evidence suggested that the participants in the 13 studies showed a gradual increase of 0.21 points in problem behavior in the baseline phase. The choice-making intervention seems to halt this increase and even sets in motion a slightly decreasing trend in problem behavior of $0.21 - 0.25 = -0.04$ points afterwards.

The standard deviations of the random effects are also given in Fig. 9. They are grouped per level: `Author:Name` refers to case-specific random effects (the case variable `Name` is nested in the study variable `Author`, see also Table 1 and Fig. 1) and `Author` refers to the study-specific random effects. We see that for all regression coefficients, the between-case variability (level 2) is higher than the between-study variability (level 3). Note again that the residual variance at the first level is close to 1 due to the analysis being performed on standardized scores. The negative correlations between the baseline intercept and the coefficients related to the effect of the treatment (`Phase` and `Time:Phase`) suggest that the more problem behavior is observed in the

Fixed effects

```
                 Estimate Std. Error        KR df    t value        p value
(Intercept)     4.7429169 0.82450619  12.441185   5.752433  7.931576e-05
Time            0.2086992 0.06215107  10.587016   3.357935  6.724094e-03
Phase1         -3.5998238 0.86605557  14.231836  -4.156574  9.374635e-04
Time:Phase1    -0.2535972 0.06382737   8.196092  -3.973172  3.905781e-03
```

Random effects

```
Groups       Name        Std.Dev. Corr
Author:Name  (Intercept) 3.972152
             Time        0.278639  1.000
             Phase1      4.103607 -0.944 -0.944
             Time:Phase1 0.251824 -0.975 -0.975  0.847
Author       (Intercept) 1.208260
             Time        0.094924  0.895
             Phase1      1.293893 -0.747 -0.371
             Time:Phase1 0.108740 -0.864 -0.998  0.310
Residual                 1.006236
```

Fig. 9 Three-level regression results tables. The three-level model was estimated based on standardized scores
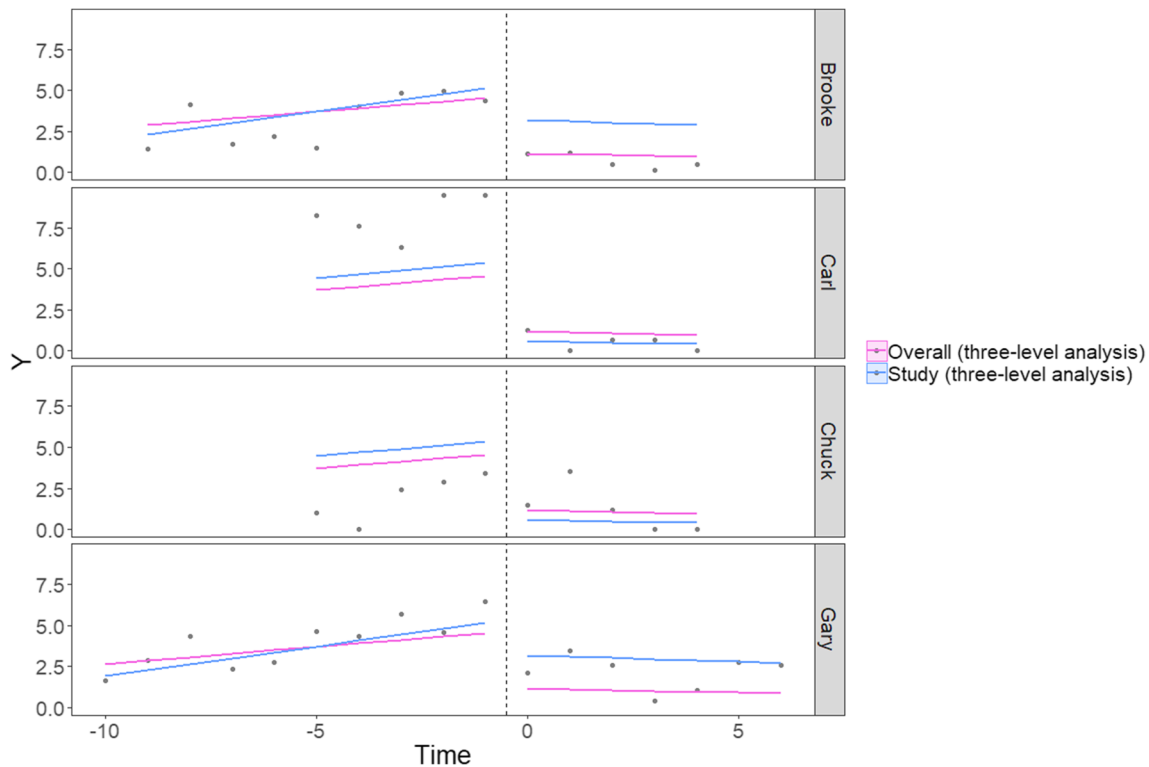
Fig. 10 Three-level regression results plot. The three-level model was estimated based on standardized scores. Brooke and Gary are participants belonging to the study by (Romaniuk et al. 2002). Carl and Chuck belong to the study by (Moes, 1998)

baseline phase, the more negative the immediate effect and the effect on the time trend and hence the more favorable the effect of the choice-making intervention is.

The results of the three-level analysis are plotted in Fig. 10 for four participants (although MultiSCED allows for creating the graphs for any number of cases and/or studies). Brooke and Gary belong to the study by Romaniuk et al. (2002), while Carl and Chuck belong to the study by Moes (1998). The blue lines represent the overall average regression lines across studies and are identical in each of the cases' graphs. The trajectory of the averages is based on the fixed effects' estimates from the table in Fig. 9: $Y = 4.74 + 0.21 \cdot \text{Time}$ in the baseline phase and $Y = (4.74 - 3.60) + (0.21 - 0.25) \cdot \text{Time} = 1.14 - 0.04 \cdot \text{Time}$ in the treatment phase. The pink lines represent the study-specific averages. They are identical for Brooke and Gary on the one hand and for Carl and Chuck on the other hand because these cases belong to the same study. Note how in the treatment phase of Brooke and Gary, the trajectory of the Romaniuk et al. (2002) study is quite different from the overall trajectory. The blue line lies considerably higher than the pink line, indicating that in this specific study, the participants showed on average more problem behavior at the start of the treatment phase than in other studies. The case-specific averages obtained from the three-level analysis are not shown in Fig. 10, but MultiSCED provides

options to plot these in the same figure, as well as the case-specific averages from the one-level analysis and the case-and study-specific averages from the two-level analysis. This allows for a thorough visual comparison of the results from all three types of analyses.

Adding moderator variables

We can extend our basic linear time trend model to verify whether other variables moderate the problem behavior outcomes. To include `Gender` and `Age` as moderator variables, we update our three-level model by modeling both variables and their interaction with `Phase` as fixed effects:

$$
\begin{aligned}
Y_{ijk} = {} & \beta_{0jk} + \beta_{1jk}\text{Time}_{ijk} + \beta_{2jk}\text{Phase}_{ijk} + \beta_{3jk} \\
& \times \left(\text{Time}_{ijk} \times \text{Phase}_{ijk}\right) \\
& + \beta_{4jk}\text{Gender}_{ijk} + \beta_{5jk}\text{Age}_{ijk} \\
& + \beta_{6jk}\left(\text{Gender}_{ijk} \times \text{Phase}_{ijk}\right) + \beta_{7jk} \\
& \times \left(\text{Age}_{ijk} \times \text{Phase}_{ijk}\right) + e_{ijk} \qquad (13)
\end{aligned}
$$

The coefficients β_{0jk}, β_{1jk}, β_{2jk} and β_{3jk} are again decomposed into a fixed and two random effects like in Eqs. 11 and 12. The other coefficients related to age and gender are considered fixed. Because standardization of the

Fixed effects

```
                 Estimate Std. Error      KR df    t value      p value
(Intercept)     4.83407017 0.88692127 15.465567   5.4503938 6.017284e-05
Time            0.19511020 0.05964516 10.167868   3.2711824 8.234355e-03
Phase1         -3.49001051 1.04476242 22.847555  -3.3404824 2.857174e-03
Gender1         0.56518570 0.56135027 12.777787   1.0068325 3.327018e-01
Age            -0.02975812 0.04364005 16.579272  -0.6818992 5.047182e-01
Time:Phase1    -0.24502566 0.06390131  8.868939  -3.8344385 4.111001e-03
Phase1:Gender1 -0.72010602 0.82562146 18.379112  -0.8721988 3.943519e-01
Phase1:Age      0.01638682 0.05842243 26.955503   0.2804885 7.812443e-01
```

Random effects

```
 Groups       Name        Std.Dev. Corr
 Author:Name (Intercept) 3.973482
             Time        0.263976  1.000
             Phase1      4.107851 -0.941 -0.941
             Time:Phase1 0.241255 -0.970 -0.970  0.832
 Author      (Intercept) 0.899803
             Time        0.088352  0.785
             Phase1      1.078611 -0.555  0.080
             Time:Phase1 0.105785 -0.739 -0.997 -0.150
 Residual                1.006676
```

Fig. 11 Three-level regression including moderator variables Age and Gender results table. The three-level model was estimated based on standardized scores

data is done based on the one-level model, whereas `Gender` and `Age` are characteristics of the units at the second level (i.e., the cases), MultiSCED still uses an adjusted model without moderators when the standardization option is checked.

The gender variable codes males as 0 and females as 1 (as described in Table 1). The coefficient named `Gender1` in the output (Fig. 11) refers to the effect of belonging to the group coded with '1' (i.e., the female participants) rather than the other group (i.e., the male participants). The `Gender1` coefficient estimate indicates that females score on average 0.55 [$t (12.78) = 1.01$, $p = .33$] points higher on problem behavior than males in the baseline phase, but this gender effect is not statistically significant. The `Phase1` estimate indicates that the choice intervention decreases problem behavior by -3.49 [$t (22.85) = -3.34$, $p < 0.01$] points. For female participants, the decrease in problem behavior is on average $-3.49 - 0.72 = -4.21$ [$t (18.38) = -0.87$, $p = 0.39$] points, although this effect is again not statistically significant. The estimates for the effect of `Age` and its impact on the treatment effect (Phase × Age) are small and appear to be non-significant.

Discussion

Relevance and implications

In their review, Manolov and Moeyaert (2017) synthesize several action points in order to decrease the gap between methodological advances and actual research practice when it comes to data-analysis of SCED data. MultiSCED provides an answer to many of those when it comes to multilevel modeling of SCED data.

First of all, the application is freely available and easy to use because of its point-and-click user interface. Using MulitSCED does not even require installing R. The application comes with an in-depth user guide (available at https://kuleuven.app.box.com/v/MultiSCEDUserGuide), featuring the same illustration as in this article using the Shogren et al. (2004) dataset, but described in much more extensive detail. The user guide is meant to be a complete tutorial, describing step by step how to prepare SCED datasets, how to upload them in MultiSCED, how to go through the application and how to interpret the results. It references the relevant methodological research

on multilevel modeling of SCED data and the underlying R packages used.

For applied researchers wanting to use multilevel modeling to analyze their SCED data, hands-on tutorials on how to get started with software tools are emerging rather slowly. Nagler, Rindskopf, and Shadish (2008) published a handbook on how to use HLM to fit multilevel models specifically for SCED data. Other works provide some pieces of code, like Baek and Ferron (2013), Ferron et al. (2009) and Moeyaert et al. (2014) for SAS. (Valentine et al., 2016) provide a tutorial for the scdhlm package Pustejovsky (2016) in R, and some R code and illustrations are found in the review of Manolov and Moeyaert (2017), as well as in the conference proceedings of Rodabaugh and Moeyaert (2017) and in the Rindskopf and Ferron (2014) book chapter. We believe that MultiSCED, through this article and the MultiSCED user guide, complements these works by illustrating the R functionality without the barrier that comes with getting to know the R syntax. Furthermore, the user guide in particular is a stand-alone document and thoroughly describes each step taken, from preparing and uploading the data file to interpreting and comparing results across models.

Another asset of MultiSCED that helps to bridge the gap is the fact that it relies heavily on graphs and visuals to represent the results. The graphs in MultiSCED are built combining tools from the popular plotting package ggplot2 (Wickham, 2009), the multilevel modeling package lme4 (Bates et al., 2015), and the data manipulation package plyr (Wickham, 2011), all made available without the need for advanced R programming expertise. Results from the one-, two-, and three-level analyses can easily be combined in one plot by using the MultiSCED customization options. Users can download and use their customized plots for reporting their results in their works.

An important option embedded within MultiSCED is the possibility to standardize the data, which is particularly useful in the three-level analysis when summarizing across studies using different measuring instruments or scales. The corresponding regression coefficients serve as standardized effect sizes for single-case data and are easily interpretable without the context of the specific SCED study they were obtained from. With the standardization option, we aim to familiarize users with standardized SCED effect sizes and criteria for standardized effect sizes for SCEDs similar to Cohen's guidelines (Cohen, 1988) for effect sizes in group-comparison designs.

Limitations and future research

MultiSCED does not offer complicated functionality in terms of modeling options (e.g., specific covariance structures other than the unstructured type or generalized linear mixed models) nor does it support complex data and design structures (e.g., ABAB designs or datasets with multiple outcomes). Applied researchers more familiar with multilevel modeling and wanting to explore such more advanced features and models further will undoubtedly find their way directly to R or to any of the dedicated software packages and tools mentioned in the introduction, guided by the references we provided throughout this paper. In order to facilitate the transition to command-line R, MultiSCED provides the R formula syntax when building regression models in the user interface.

We have deliberately not given much attention to the underlying R code on the server side of MultiSCED, because with this article and with the user guide, we first of all want to address novices who prefer to avoid the R syntax. However, it could be useful to follow-up on this article and the user guide with a more technical tutorial on how to use the lme4 package to model SCED data and how to obtain the plots shown in MultiSCED.

We are not presenting any new techniques with MultiSCED, nor is multilevel modeling the one-size-fits-all solution when it comes to quantitative data analysis of SCEDs, but precisely because of the myriad of techniques proposed, it is important that we provide tools for applied researchers so that they are able to choose the appropriate method for their data and apply it in a correct way. Specifically for R, a challenge for the future is to combine the individual pieces of software that already exist into one comprehensible tool or package. Merging MultiSCED with the R code provided by Manolov and Moeyaert (2017), the functionality from the SCDA plug-in for R commander (Bulté & Onghena, 2013), the SSDforR package (Auerbach & Schudrich, 2013), the scdhlm package (Pustejovsky, 2016) and other existing Shiny applications like the application at https://manolov.shinyapps.io/SeveralAB (Manolov & Rochat, 2015; Manolov & Solanas, 2018), and streamlining the end result might be very useful to bring applied researchers to R for analyzing their SCED data. For an overview of tools in R and other software options, we refer to the list of single-case data analysis software tools at https://osf.io/sdv4m/.

The meta-analytic dataset by Shogren et al. (2004) is used in this paper as a representative example of a meta-analysis of SCEDs. According to the review by Shadish and Sullivan (2011), the primary studies in this meta-analytic dataset are slightly smaller than average with 2.31 cases per study and 13.27 measurements per case on average. With 13 studies, the number of included studies is also rather low (Moeyaert et al. 2013). When applying the multilevel techniques as we did with MultiSCED, sample size recommendations based on simulation studies are usually not strict when it comes to the validity of the overall fixed effects. However, variance estimations are

often biased for small sample sizes (Moeyaert et al., 2014; Ferron et al., 2009; Owens & Ferron, 2012). Moeyaert et al. (2013) have empirically validated a two-level model for standardized SCED data and recommend to have at least 20 measurements per case, and even more (at least 40) in order to obtain reliable variance estimations. We therefore repeat the warnings found in previous simulation studies to interpret the variance estimations based on the Shogren et al. (2004) data with caution.

Conclusions

In this article, we presented MultiSCED, a free R Shiny application available at http://www.single-case.com/MultiSCED and aimed at applied SCED researchers wanting to get started with multilevel modeling to analyze their SCED data obtained from multiple cases or multiple studies. The application provides R functionality within a point-and-click user interface, thus lowering the threshold that inexperienced R users might face. Through a re-analysis of a real meta-analytical dataset, we illustrated the use of MultiSCED step by step and we showed how to interpret the numerical multilevel results graphically. With this tutorial we try to contribute to bridging the gap between methodological research and applied research practices and to make advanced SCED data analysis methods more accessible for a wider, applied audience.

References

Auerbach, C., & Schudrich, W. Z. (2013). SSD for R: A comprehensive statistical package to analyze single-system data. *Research on Social Work Practice*, 23(3), 346–353. https://doi.org/10.1177/1049731513477213

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baek, E. K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across-participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45(1), 65–74. https://doi.org/10.3758/s13428-012-0231-z

Baek, E. K. et al. (2016). Using visual analysis to evaluate and refine multilevel models of single-case studies. *Journal of Special Education*, 50(1), 18–26. https://doi.org/10.1177/0022466914565367

Bates, D. (2006). lmer, p-values and all that. https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bolker, B. et al. (2018). GLMM FAQ. http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Bulté, I., & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods*, 12(2), 450–478. https://doi.org/10.1017/CBO9781107415324.004

Chang, W. et al. (2017). Shiny: Web application framework for R. https://cran.r-project.org/package=shiny

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Erlbaum.

Declercq, L. et al. (2018). Analysis of single-case experimental count data using the linear mixed effects model: A simulation study. Behavior Research Methods. https://doi.org/10.3758/s13428-018-1091-y

Dibley, S., & Lim, L. (1999). Providing choice making opportunities within and between daily school routines. *Journal of Behavioral Education*, 9(2), 117–132. https://doi.org/10.1023/A:1022888917128

Ferron, J. M. et al. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, 41(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods*, 42(4), 930–943. https://doi.org/10.3758/BRM.42.4.930

Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, 54(1-3), 137–154. https://doi.org/10.1016/S0376-6357(01)00155-3

Jamshidi, L. et al. (2018). Review of single-subject experimental design meta-analyses and reviews: 1985-2015. Submitted for publication.

Joo, S. H. et al. (2017). Approaches for specifying the level-1 error structure when synthesizing single-case data. *Journal of Experimental Education*, 0973, 1–20. https://doi.org/10.1080/00220973.2017.1409181

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53(3), 983–997. https://doi.org/10.2307/2533558

Kern, L. et al. (2001). Choice of task sequence to reduce problem behaviors. *Journal of Positive Behavior Interventions*, 3(1), 3–10. https://doi.org/10.1177/109830070100300102. http://search.ebscohost.com/login.aspx?direct=true%7B%5C&%7Ddb=psyh%7B%5C&%7DAN=2001-14259-001%7B%5C&%7Dsite=ehost-live

Littell, R. C. et al. (2007). SAS for mixed models. 2nd ed., 2. Cary: SAS Institute.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, 51(2), 201–215. https://doi.org/10.1016/j.jsp.2012.12.005

Manolov, R., & Rochat, L. (2015). Further developments in summarising and meta-analysing single-case data: An illustration with neurobehavioural interventions in acquired brain injury. *Neuropsychological Rehabilitation*, 25(5), 637–662. https://doi.org/10.1080/09602011.2015.1064452

Manolov, R., & Moeyaert, M. (2017). How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behavior Modification*, 41(2), 179–228. https://doi.org/10.1177/0145445516664307

Manolov, R., & Solanas, A. (2018). Quantifying differences between conditions in single-case designs: Possible analysis and meta-analysis. *Developmental Neurorehabilitation*, *21*(4), 238–252. https://doi.org/10.3109/17518423.2015.1100688

Moes, D. R. (1998). Integrating choice-making opportunities within teacher-assigned academic tasks to facilitate the performance of children with autism. *Journal of the Association for Persons with Severe Handicaps*, *23*(4), 319–328.

Moeyaert, M. et al. (2013). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research*, *48*(5), 719–748. https://doi.org/10.1080/00273171.2013.816621

Moeyaert, M. et al. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, *52*(2), 191–211. https://doi.org/10.1016/j.jsp.2013.11.003

Moeyaert, M. et al. (2015). Estimating intervention effects across different types of single-subject experimental designs: Empirical illustration. *School Psychology Quarterly*, *30*(1), 50–63. https://doi.org/10.1037/spq0000068

Moeyaert, M. et al. (2017). Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation. *Psychological Methods*, *22*(4), 760–778. https://doi.org/10.1037/met0000136

Nagler, E., Rindskopf, D. M., & Shadish, W. R. (2008). Analyzing data from small N designs using multilevel models: A procedural handbook. Unpublished manuscript.

Onghena, P., & Edgington, E. S. (2005). Customization of pain treatments: Single-case design and analysis. *The Clinical Journal of Pain*, *21*(1), 56–68. https://doi.org/10.1097/00002508-200501000-00007

Owens, C. M., & Ferron, J. M. (2012). Synthesizing single-case studies: A Monte Carlo examination of a three-level meta-analytic model. *Behavior Research Methods*, *44*(3), 795–805. https://doi.org/10.3758/s13428-011-0180-y

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T.R. Kratochwill, J.R. Levin, & N. J. Hillsdale (Eds.) *Single-case research design and analysis: New directions for psychology and education*. Erlbaum. Chap. 2, (pp. 15–38).

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, *65*(5), 717–741. https://doi.org/10.1177/0013164405278558

Pinheiro, J. et al. (2018). nlme: Linear and nonlinear mixed effects models. R package version 3. pp. 1-137. https://CRAN.R-project.org/package=nlme

Pustejovsky, J. E. (2016). *scdhlm: Estimating hierarchical linear models for single-case designs. R package version 0.3*. Austin: University of Texas. http://github.com/jepusto/scdhlm

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria. http://www.r-project.org/

Rasbash, J. et al. (2009). A user's guide to MLwiN, Version 2.10, pp. 1–296. http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/mlwin-userman-09.pdf

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2013). HLM 7.01 for Windows [computer software]. Skokie, IL.

Rindskopf, D. M., & Ferron, J. M. (2014). Using multilevel models to analyze single-case design data. In R.T. Kratochwill, & J.R. Levin (Eds.) *Single-case intervention research: Methodological and statistical advances*. School psychology series. Washington, DC, US: American Psychological Association, pp. 221–246. https://doi.org/10.1037/14376-008

Rodabaugh, E., & Moeyaert, M. (2017). Multilevel modeling of single-case data: An introduction and tutorial for the applied researcher. In *NERA Conference Proceedings 2017 8.*

Romaniuk, C. et al. (2002). The influence of activity choice on problem behaviors maintained by escape versus attention. *Journal of Applied Behavior Analysis*, *35*(4), 349–362. https://doi.org/10.1901/jaba.2002.35-349. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1284398/

Shadish, W. R. (2014). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, *23*(2), 139–146. https://doi.org/10.1177/0963721414524773

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*(4), 971–980. https://doi.org/10.3758/s13428-011-0111-y

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*(3), 188–196. https://doi.org/10.1080/17489530802581603. http://search.ebscohost.com/login.aspx?direct=true%7B%5C&%7Dlang=zh-tw%7B%5C&%7Dsite=ehost-live

Shadish, W. R., Kyse, E. N., & Rindskopf, D. M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods*, *18*(3), 385–405. https://doi.org/10.1037/a0032964

Shogren, K. A., Faggella-Luby, M. N., Bae, S. J., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions*, *6*(4), 228–237. https://doi.org/10.1177/10983007040060040401

Ugille, M. et al. (2012). Multilevel meta-analysis of single-subject experimental designs: A simulation study. *Behavior Research Methods*, *44*(4), 1244–1254. https://doi.org/10.3758/s13428-012-0213-1

Valentine, J. C. et al. (2016). Between-case standardized mean difference effect sizes for single-case designs: A primer and tutorial using the scdhlm web application. Oslo, Norway: The Campbell Collaboration. https://doi.org/10.4073/cmdp.2016.1. https://campbellcollaboration.org

Van den Noortgate, W. et al. (2014). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*(3), 325–346. https://doi.org/10.1521/scpq.18.3.325.22577

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 1–10. https://doi.org/10.3758/BF03195492

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 142–151. https://doi.org/10.1080/17489530802505362

Welham, S. J., & Thompson, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *59*(3), 701–714. https://doi.org/10.2307/2346019

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer. ISBN: 978-0-387-98140-6. http://ggplot2.org.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. http://www.jstatsoft.org/v40/i01/