KU LEUVEN

FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

Onderzoekseenheden
Methodologie van het Pedagogisch Onderzoek en Gezins-en Orthopedagogiek

PSYCHOSOCIAL INTERVENTIONS FOR REDUCING VOCAL CHALLENGING BEHAVIOUR IN PERSONS WITH AUTISTIC DISORDER:

A multilevel meta-analysis of single-case experiments

Masterproef aangeboden tot het verkrijgen van de graad van Master of Science in de Pedagogische Wetenschappen

Door
Lien Vanderkerken

promotor: Prof. Dr. Patrick Onghena
copromotoren: Prof. Dr. Bea Maes
Dr. Mieke Heyvaert

2013

KU LEUVEN

FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN

Onderzoekseenheden
Methodologie van het Pedagogisch Onderzoek en Gezins-en Orthopedagogiek

PSYCHOSOCIAL INTERVENTIONS FOR REDUCING VOCAL CHALLENGING BEHAVIOUR IN PERSONS WITH AUTISTIC DISORDER:

A multilevel meta-analysis of single-case experiments

Masterproef aangeboden tot het verkrijgen van de graad van Master of Science in de Pedagogische Wetenschappen

Door
Lien Vanderkerken

promotor: Prof. Dr. Patrick Onghena
copromotoren: Prof. Dr. Bea Maes
Dr. Mieke Heyvaert

2013

Lien, Vanderkerken, Psychosocial interventions for reducing vocal challenging behaviour in persons with autistic disorder: A multilevel meta-analysis of single-case experiments.
Masterproef aangeboden tot het verkrijgen van de graad van Master in de Pedagogische Wetenschappen
Examenperiode: Juni 2013
Promotor: Prof. Dr. Patrick Onghena
Copromotoren: Prof. Dr. Bea Maes, Dr. Mieke Heyvaert

Vocal challenging behaviour (VCB) forms a common problem in individuals with autistic disorder. Examples of VCB are vocal stereotypy, inappropriate vocalisations, verbal aggression, and screaming. Since VCB is associated with negative outcomes for the individual and for his or her environment, it is important to know how to manage this type of CB.

To evaluate the effectiveness of several psychosocial interventions applied to decrease VCB in individuals with autistic disorder, we conducted a meta-analysis of single-case experiments (SCEs).

We did a systematic search for SCEs consulting 4 databases and 23 relevant journals. Subsequently, we coded the retained articles for several variables at the level of the participant, at the level of the intervention and the intervention context, and at the level of the study. In addition, we retrieved the raw data from the graphs presented in the articles.

The SCE data were combined using multilevel models in order to answer the following four research questions: (1) What is the overall effect of psychosocial interventions for VCB in individuals with autistic disorder? (2) Are there differences between studies? (3) Are there differences between participants? (4) What characteristics at the level of the participant, at the level of the intervention and the intervention context, and at the level of the study have a moderating effect on the intervention effect?

We estimated and tested the overall intercept and the overall effect of psychosocial interventions for VCB and estimated and tested the (co)variances between studies and participants. Furthermore, based on a priori considerations and guided by the data (i.e., significance tests, estimated moderator effects, and fit indices), we built the definitive three-level model containing seven potential moderating variables. Subsequently, we examined this three-level model in detail. In addition, we conducted a sensitivity analysis and a publication bias analysis.

The overall treatment effect was large and statistically significant (cf. research question 1). However, this effect varied significantly over the included studies and participants (cf. research questions 2 and 3). VCB type and intervention type significantly moderated the treatment effect, with, on average, the largest effects for interventions used to reduce VCB including stereotypical VCB and for interventions containing both antecedent and consequence components. Age, gender, primary treatment setting, publication year, and study quality did not significantly moderate the intervention effect (cf. research question 4). According to the sensitivity analysis, these results were quite robust. Furthermore, based on a visual inspection of the funnel plot and a regression test for funnel plot asymmetry, we concluded that the present meta-analysis does not suffer much from publication bias effects, although publication bias cannot be excluded absolutely. The quantitative combination of multiple SCEs offered the opportunity to explore the generalizability of separate SCEs and to obtain information about the overall effect, about specific cases, and about the effect of moderating variables on the treatment effect.

Woord van dank

Een masterproef uitwerken doe je niet op één dag en bovendien niet alleen. Daarom wil ik ook graag alle personen die een belangrijke rol hebben gespeeld bij de totstandkoming van dit resultaat, bedanken.

Allereerst wil ik mijn promotor Prof. Dr. Patrick Onghena bedanken voor zijn enthousiasme, de inspirerende en constructieve feedback, en het uitdagende motto: “Alles kan altijd beter!”.
Vervolgens wil ik graag Prof. Dr. Bea Maes danken voor haar kritische vragen en waardevolle opmerkingen.
Ook ben ik veel dank verschuldigd aan Dr. Mieke Heyvaert voor haar snelle en uitgebreide feedback, de vele leerkansen, en de talrijke nuttige tips.
Daarenboven verdienen de medewerkers van de Pbib een woord van dank voor het mee zoeken naar jaargangen van tijdschriften, het bestellen van artikels, en de fijne werksfeer in de bibliotheek.

Verder wil ik ook mijn ouders danken omdat ze me de kans gaven om te studeren en mijn interesses te volgen.

Ten slotte bedank ik ook graag mijn familie en vrienden voor de bemoedigende berichtjes en gesprekken, de ontspannende momenten op kot, in de Alma, thuis, en in de trein, en vooral voor hun begrip voor mijn soms langdurige isolatie.

Aan jullie allen, dank u wel!

Toelichting aanpak en eigen inbreng

De voorliggende masterproef betreft een meta-analyse van single-case experimenten over psychosociale interventies voor vocaal probleemgedrag bij personen met een autistische stoornis.

Gebaseerd op de studie van Campbell (2003), een meta-analyse van single-case experimenten over gedragstherapeutische interventies voor probleemgedrag bij mensen met een autistische stoornis, voerde ik tijdens het eerste masterjaar een verkennende zoektocht uit naar geïndividualiseerd onderzoek over gedragstherapeutische interventies voor probleemgedrag bij mensen met een autistische stoornis. Mede omwille van de zeer grote dataset werd het onderwerp vervolgens aangepast. We kozen ervoor het onderzoek te verfijnen tot vocaal probleemgedrag en ook sociaal-contextuele interventies te onderzoeken. Een nieuw zoekproces werd uitgevoerd.

In eerste instantie was het de bedoeling om zowel single-case experimenten als kwalitatieve case studies over psychosociale interventies voor vocaal probleemgedrag bij mensen met een autistische stoornis te combineren in een mixed methods meta-onderzoek. Vermits we geen kwalitatieve case studies over het besproken onderwerp vonden, beperkt deze masterproef zich tot een kwantitatieve synthese van single-case experimenten. Desalniettemin bestudeerde ik ook literatuur met betrekking tot case studies, kwalitatieve synthesetechnieken, mixed methods onderzoek, en mixed methods meta-onderzoek.

Ikzelf voerde het systematische zoekproces naar geïndividualiseerd onderzoek over vocaal probleemgedrag bij mensen met een autistische stoornis uit. Dr. Mieke Heyvaert repliceerde, als tweede onderzoeker, een deel van dit zoekproces. Ook controleerde zij de door mij geselecteerde artikels aan de hand van de inclusie- en exclusiecriteria. Hierna codeerde ik alle definitief weerhouden artikels. Dr. Mieke Heyvaert herhaalde, ter controle, het volledige coderingsproces. Vervolgens extraheerde ik de ruwe data uit de grafieken en voerde ik de data-analyse uit, aan de hand van SAS-codes en R-codes die mij door Dr. Mieke Heyvaert ter beschikking werden gesteld. Ten slotte schreef ik onderstaand artikel.

Omdat deze masterproef deels voortbouwt op het artikel van Campbell (2003) dat gepubliceerd werd in *Research in Developmental Disabilities*, kozen we voor een Engelstalige masterproef, in artikelvorm. *Research in Developmental Disabilities* was dan ook het beoogde tijdschrift. Daarenboven bood een artikelvorm me de kans om het onderzoeksproces, de verantwoording, en de resultaten op een beknopte manier weer te geven, wat zeker en vast ook een leerkans inhield. Verder zag ik deze masterproef als een gelegenheid om Engelstalige academische schrijfvaardigheden te ontwikkelen.

Contents

Psychosocial Interventions for Reducing Vocal Challenging Behaviour in Persons with Autistic Disorder:
A Multilevel Meta-analysis of Single-case Experiments

Lien Vanderkerken, Mieke Heyvaert, Bea Maes, and Patrick Onghena
KU Leuven, Belgium

Abstract

Vocal challenging behaviour (VCB) forms a common problem in individuals with autistic disorder. Since VCB is associated with negative outcomes for the individual and his or her environment, it is important to know how to manage this type of CB. To evaluate the effectiveness of several psychosocial interventions applied to decrease VCB in individuals with autistic disorder, we conducted a meta-analysis of single-case experiments (SCEs). Fifty-two SCEs, including 74 participants, were combined using a multilevel meta-analysis. The overall treatment effect was large and statistically significant. However, the effect varied significantly over the included studies and participants. Examining this variance, evidence was found for a moderator effect of VCB type and intervention type, with, on average, the largest effects for interventions used to reduce VCB including stereotypical VCB and for interventions containing both antecedent and consequence components. Age, gender, primary treatment setting, publication year, and study quality did not significantly moderate the intervention effect.

Keywords: meta-analysis, systematic review, single-case, interventions, vocal challenging behaviour, vocal problem behaviour, vocal stereotypy, autism, autistic disorder

1. Introduction

Individuals with autistic disorder present an impaired or abnormal development in social interactions, communication, and/or symbolic play that is manifest before the age of three years. They have qualitative deficits in social skills and communication and show repetitive, stereotyped patterns of behaviour, interests, and activities (American Psychiatric Association, 2000; World Health Organization, 2013). Especially these deficits in social skills, language and communication, and the rigid behaviours make them vulnerable to develop challenging behaviours (CB; Buschbacher & Fox, 2003; Howlin, 1998). Besides, persons with autistic disorder often present intellectual disabilities (ID) and/or co-occurring psychopathologies, factors that increase the risk of CB (Helverschou, Bakken, & Martinsen, 2011; Holden & Gitlesen, 2009; Matson & Shoemaker, 2009). A common form of CB in individuals with autistic disorder is *vocal challenging behaviour (VCB).* Examples are vocal stereotypy, verbal aggression, inappropriate vocalisations, and screaming (Cohen, Yoo, Goodwin, & Moskowitz, 2011; Healy & Leader, 2011; Lequia, Machalicek, & Rispoli, 2012).

The literature about the causal and maintaining factors of VCB is very limited and is often restricted to the operant and the homeostatic function of (vocal) stereotypy (e.g., Cunningham & Schreibman, 2008; Hodgetts, Magill-Evans, & Misiaszek, 2011). Sometimes, the neurobiology of (vocal) stereotypy is discussed (e.g., Langen, Durston, Kas, van Engeland, & Staal, 2011; Miguel, Clark, Tereshko, & Ahearn, 2009). Occasionally, the impaired perspective taking, imagination, and language development are mentioned as contributing factors to VCB in persons with autistic disorder (e.g., Hetzroni & Tannous, 2004; Norris & Dattilo, 1999). Hence, for a comprehensive understanding of the onset and maintenance of VCB in individuals with autistic disorder, we ground on the biopsychosocial model on CB of Došen, Gardner, Griffiths, King, and Lapointe (2007). According to this integrative model, several biological (e.g., pain, sleep irregularities), psychological (e.g., experiences of distress, anxiety) and environmental (e.g., lack of structure, attention) factors are involved in the onset of CB. They can function as instigating, processing, and maintaining conditions. Similarly, these factors contribute to psychiatric disorders (e.g., Tourette's syndrome), that can also interact with CB. This biopsychosocial paradigm has to be placed in a developmental perspective since a person's cognitive, emotional, social, and personality development and his or her functional skills influence the way he or she experiences and reacts to events. Accordingly, a school-age child with autistic disorder who has no functional speech can feel very distressed during unexpected changes and can start screaming.

Although VCB does not immediately cause physical harm like for instance self-injurious behaviour (cf. Emerson, 2001), it is associated with negative outcomes for the individual concerned and for his or her environment. VCB may impede a person from learning and social interactions and can be socially stigmatising (e.g., Keeling, Smith Myles, Gagnon, & Simpson, 2003; MacDonald et al., 2007). It may hinder peers (e.g., Liu-Gitz & Banda, 2010) and can involve stress for parents, teachers, therapists, and staff (cf. Hastings, 2002, 2005; Tomanik, Harris, & Hawkins, 2004). VCB thus affects the quality of life of the individual and of his or her environment (cf. Schalock, 2004). Consequently, it is important to know how to manage this type of CB.

Managing VCB forms a specific challenge. To start, VCB cannot be physically stopped or interrupted (Lanovaz & Sladeczek, 2012). In addition, several authors report the difficulty of assessing and treating automatically maintained vocal stereotypy (e.g., Athens, Vollmer, Sloman, & Pipkin, 2008; Rapp, 2007). The produced sensory reinforcers are rather inaccessible for parents, teachers, or therapists who want to diminish the behaviour (Vollmer, 1994). Furthermore, individuals with autistic disorder have sometimes very limited communicative skills and language capacities. Hence, teaching alternative verbal responses can be difficult (Franco et al., 2009; Hutchins & Prelock, n.d.). Besides, it could happen that interventions intended to reduce VCB simultaneously decrease positive behaviour (e.g., appropriate vocalisations), which is not desirable (Lanovaz & Sladeczek, 2012).

Today, several *psychosocial interventions* like for instance differential reinforcement (DR), response interruption and redirection (RIRD), and augmentative and alternative communication (AAC) are applied to reduce VCB in persons with autistic disorder (e.g., Franco et al., 2009; Lanovaz & Sladeczek, 2012). These interventions have been evaluated in multiple primary-level studies. Except from some group studies including also persons with other autism spectrum disorders (ASD; e.g., Gabriels et al., 2012), to our knowledge, all the published primary-level articles on this topic concern *single-case experiments (SCEs).*

In SCEs, causal relationships between independent and outcome variables are studied within one entity or case by manipulating the independent variables and conducting repeated measurements of the outcome variables under the different treatment conditions (Barlow, Nock, & Hersen, 2009; Kazdin, 2011). Although experimental group-comparison studies are often seen as the gold standard in scientific enquiry (Grossman & Mackenzie, 2005), in the research domain of interventions for CB among persons with developmental disabilities (DD), SCEs have several advantages over group-comparison designs (Matson, Turygin, Beighley, & Matson, 2012). To start, SCEs need only one, or a small number of cases. In

addition, while group-comparison studies aggregate data for a group of participants, SCEs focus on the individual and offer an in-depth insight into the behaviour of a single case. Moreover, by conducting many repeated observations, the evolution of the behaviour can be studied in detail. That way, SCEs offer more opportunities for valid person-specific decisions. Finally, SCEs provide a cost-effective approach and are quite easy to set up (Horner et al., 2005; Van den Noortgate & Onghena, 2007; Zhan & Ottenbacher, 2001). Although SCE research can reveal valid conclusions about the studied cases, replication and aggregation are needed to explore and enlarge the generalizability of SCE results (Iwakabe & Gazzola, 2009; Van den Noortgate & Onghena, 2007). *Meta-analysis*, the statistical aggregation of individual study results (Glass, 1976), offers such an opportunity.

So far, various SCE meta-analyses on interventions for CB among persons with DD have been published. These meta-analyses show a lot of variation on the studied participants, CB, and interventions (e.g., Campbell, 2003; Heyvaert, Maes, Van den Noortgate, Kuppens, & Onghena, 2012; Shogren, Faggella-Luby, Bae, & Wehmeyer, 2004). Campbell (2003) studied behavioural interventions for various types of CB in persons with autistic disorder. Ma (2009) conducted an SCE meta-analysis on psychosocial interventions in persons with autism to demonstrate the percentage of datapoints exceeding the median of baseline phase (PEM) approach. In addition, some meta-analyses studied the effect of specific interventions on the CB of persons with ASD (Ganz et al., 2012; Kokina & Kern, 2010; Lequia et al., 2012). In line with Campbell (2003), the present study concerns individuals with autistic disorder. In addition, we will pay attention to a specific type of CB, namely VCB.

Even though multiple SCEs on psychosocial interventions for VCB in individuals with autistic disorder have been published, to date, they were not systematically aggregated in a separate review. Lanovaz and Sladeczek (2012) reviewed SCEs on behavioural interventions for vocal stereotypy in individuals with ASD. However, this study was a not a *systematic review* (cf. Higgins & Green, 2011), neither did the authors apply a *statistical meta-analysis*. Furthermore, the review of Lanovaz and Sladeczek (2012) was restricted to automatically maintained stereotypical VCB. Nevertheless, also other VCB types (e.g., socially reinforced vocal stereotypy, screaming) are often seen in this population (cf. Cunningham & Schreibman, 2008; Galiatsatos & Graff, 2003).

In order to fill these knowledge gaps, we conducted a meta-analysis of SCEs on the effectiveness of psychosocial interventions for VCB in individuals with autistic disorder. We aimed to answer the following four questions: (1) What is the overall effect of psychosocial interventions for VCB in individuals with autistic disorder? (2) Are there differences between

studies? (3) Are there differences between participants? (4) What characteristics at the level of the participant, at the level of the intervention and the intervention context, and at the level of the study have a moderating effect on the intervention effect?

2. Methods

2.1. Systematic and Comprehensive Search for Evidence

2.1.1. Inclusion and exclusion criteria. We aimed at reviewing SCEs on psychosocial interventions for VCB among individuals with autistic disorder. The following inclusion and exclusion criteria were used:

1. To be retained, articles had to include at least one participant diagnosed with autistic disorder (cf. American Psychiatric Association, 2000; World Health Organization, 2013). When articles included multiple subjects, solely the persons with autistic disorder were included in the meta-analysis. Persons only reported as autistic-like or presenting autistic-like behaviour were excluded (cf. Campbell, 2003). No age-related criteria were applied.
2. Studies were selected when participants with autistic disorder exhibited VCB. Only data for VCB were included in the meta-analysis. Data for positive behaviour, data for other forms of CB, and data in which VCB was aggregated with other CB types were excluded.
3. Data on psychosocial interventions were retained. Accordingly, we included data on psychosocial interventions directed to the individual and/or the environment. Data on pharmacological interventions were excluded. To be included, participants who took medications had to take them during both baseline and treatment phases.
4. Only SCEs were selected. Studies had to offer repeated baseline and treatment datapoints and had to present raw data for each participant separately (i.e., neither mean scores, nor aggregated data for multiple subjects). Both baseline and treatment condition had to contain at least two datapoints (cf. Heyvaert et al., 2012). Articles using group-comparison designs and group studies without a display of individual participant characteristics were not selected. Also research syntheses were excluded since meta-studies (e.g., meta-analyses, systematic reviews) do not report raw data for individual participants. Further, articles that did not report primary empirical results on effects of interventions for VCB in individuals with autistic disorder (e.g., editor's pieces, opinion articles) were excluded. Experiments on VCB without an intervention to reduce this behaviour and studies limited to a functional analysis of VCB were also excluded.
5. In order to conduct a systematic review of recent research published in the last two decades, the studies had to be published in the period January 1990–December 2011.

6. Striving for a good understanding of the articles by the coders, the articles had to be written in English.

2.1.2. Systematic search process. The systematic search for SCEs was based on the search for SCEs on behavioural interventions for CB in individuals with autistic disorder of Campbell (2003). First, we systematically searched four databases: *ERIC, MedLine, PubMed*, and *Web of Science*, applying the search string: (*autism* OR *autistic*) AND (*vocal**). For *MedLine* and *PubMed*, the options *humans* and *English language* were selected. Also during searching *Web of Science*, the language option *English* was chosen. Second, we manually searched 23 relevant journals. The searched journals are presented in Appendix A.

2.1.3. Selected studies and reliability. The search for relevant manuscripts in the 4 databases and 23 journals was conducted by the first author. As a reliability check, the second author checked 25% of the search process: all titles and abstracts of the retrieved articles from one randomly selected (Haahr, 1998) database (i.e., *Web of Science*) and six randomly selected journals (i.e., *Autism Research, Behavior Modification, Focus on Autism and other Developmental Disabilities, Intellectual and Developmental Disabilities, Journal of Abnormal Child Psychology,* and *Journal of Experimental Child Psychology*) were independently screened for inclusion by the second author. Full text copies of all potentially relevant papers were retrieved. The inter-rater agreement was 100%.

Altogether, the first author selected 54 articles for inclusion: 21 articles were retained by screening the databases, and 33 additional articles were selected during the hand search of the journals. As a check for the final inclusion, the second author read the 54 full text copies. There was disagreement on the inclusion of two articles (both articles were not retrieved through *Web of Science* nor through the six randomly selected journals). After discussion between the first and the second author, it was decided to exclude both articles.

The final database included 52 articles. The articles included by searching the databases were not published in any other journal than the manually searched journals. Appendix A offers more information about the searched journals and the corresponding numbers of retained articles. Appendix B contains a reference list of the included articles.

2.2. Data Extraction

2.2.1. Case and study characteristics. Based on a reading of various SCE meta-analyses on interventions for CB among individuals with autism and/or other DD (e.g., Campbell, 2003; Harvey, Boer, Meyer, & Evans, 2009), we coded the selected articles for certain descriptive and several potential moderating variables at the level of the participant,

at the level of the intervention and the intervention context, and at the level of the study.

At the level of the participant, we coded the variables *age* and *gender*, the *VCB type* targeted by the intervention and the *function of vocal stereotypy*. Furthermore, we coded the presence or absence of an *additional language problem*, the presence or absence of an *additional disability or disorder* (e.g., sensory impairment, Tourette's syndrome), the *presence or absence and the level of an ID*, and whether the participant took *medication*. At the level of the intervention and the intervention context, we coded the *use and the type of a pretreatment functional analysis (FA)*, the applied *intervention type*, the *primary treatment setting*, the *primary intervention agent*, the *involvement of peers*, the study *design*, the *duration of treatment*, the *frequency of treatment* (the combination of the duration and the frequency can be seen as the "dose" of treatment), and the number of *treatment datapoints*. At the level of the study, the variables *publication year* and *study quality* were coded. For an overview of these variables, we refer to Appendix C.

First, we expected that psychosocial interventions would be effective in reducing VCB in individuals with autistic disorder. Second, we hypothesised that all potential moderators of Appendix C could moderate the intervention effect. Third, and in line with Campbell (2003) and Ma (2009), we expected that especially VCB type and intervention type could function as moderators. Since stereotypy is one of the core features of autistic disorder (American Psychiatric Association, 2000; World Health Organization, 2013), VCB type was divided into (a) *stereotypical VCB* (e.g., repetitive sounds, perseverative speech, vocal tics), (b) *non-stereotypical VCB* (e.g., screaming, talking-out, irrelevant speech), and (c) *combined VCB*, containing both stereotypical VCB and non-stereotypical VCB (e.g., inappropriate vocalisations and echolalia). Based on the reviews of Lanovaz and Sladeczek (2012) and Machalicek, O'Reilly, Beretvas, Sigafoos, and Lancioni (2007), intervention type was divided into (a) *antecedent interventions*, in which actions are taken prior to or independent of the occurrence of VCB (e.g., social stories, noncontingent reinforcement), (b) *consequence interventions*, in which actions depend on the presence or absence of VCB (e.g., RIRD, DR), and (c) *combined interventions*, containing both antecedent and consequence components (e.g., social stories with contingent praise). Fourth, we expected that a pretreatment FA (i.e., the use of indirect, descriptive, or experimental assessment methods in order to identify the function of CB; Tarbox et al., 2009) would lead to larger decreases in VCB. Fifth, since previous SCE meta-analyses found a positive association between the use of a pretreatment FA and publication decade, and better results for interventions including a pretreatment FA (e.g., Campbell, 2003), more recent studies were expected to have better treatment results.

The Single-Case Experimental Design (SCED) Scale (Tate et al., 2008) was used to code the variable study quality. We evaluated the methodological quality of the retrieved SCEs through the Items 2-11. Based on its development process and multiple studies on inter-rater reliability, the SCED Scale can be considered as a highly valid tool for the quality assessment of SCEs, with high levels of inter-rater reliability (Tate et al., 2008).

2.2.2. Treatment effectiveness. The results of the repeated measurements were retrieved from the primary studies. Because these raw data were offered graphically, a digitiser program was used. With the UnGraph software (Version 5.0) a grid was constructed over the graphs presented in the articles. The X (i.e., session, day, etc.) and Y (i.e., a measure of VCB) coordinates of each datapoint were determined. The study of Shadish et al. (2009) revealed that extraction of SCE data with UnGraph can be very reliable and valid when researchers are well informed and carefully monitor their actions. The numeric output for each participant was inserted in Microsoft Excel and was combined with the previously coded case-specific characteristics, in order to get the final data file.

2.2.3. Reliability of coding. The coding of case and study characteristics for the 74 participants was done by the first author. As a reliability check, the second author independently recoded all variables for all included participants. Both authors coded 2370 items. The reliability check resulted in 2350 agreements. There was a disagreement for 20 items. The inter-rater agreement, calculated by dividing the number of agreements by the number of agreements plus disagreements, was 99.2%. After discussion between the first and second author, the codes as agreed were given to the items.

2.3. Data Analysis

2.3.1. Standardisation and descriptive analysis. The selected studies utilised various measures of VCB (e.g., percentage of time intervals with VCB, frequency of VCB) and different time intervals and session lengths. Hence, the obtained data were not immediately comparable. To solve this issue, the data were standardised. Using SAS (Version 9.3; SAS Institute, 2011), we conducted a series of ordinary participant-specific regression analyses, whereby VCB was predicted by the condition (i.e., baseline or treatment). That way, the root mean squared errors were estimated. Subsequently, the raw data of each participant were divided by the participant's root mean squared error in order to get standardised data (cf. Van den Noortgate & Onghena, 2008).

Furthermore, before conducting the meta-analysis, we carried out a descriptive analysis to get more insight into the data (cf. Snijders & Bosker, 2012). The obtained

frequencies, means, standard deviations, ranges, and correlations of possible moderators (and of descriptive variables) are presented in Appendix C.

2.3.2. Multilevel meta-analysis of single-case experiments. Our data are hierarchically structured and show a three-level configuration (cf. Van den Noortgate & Onghena, 2003a, 2008). Repeated measurements (first level) are clustered within participants (second level), who are clustered within studies (third level). As a result, the scores of one participant can be more similar than the scores of different participants. The scores of participants from the same study can also be more alike than those of participants belonging to different studies. To take into account these potential dependencies, we used *multilevel models* (cf. Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) to conduct the meta-analysis of SCE data (cf. Van den Noortgate & Onghena, 2003a, 2003b, 2003c, 2007, 2008).

A multilevel model is formed by a set of different-level regression equations. The coefficients of the higher-level equations serve as predictors for the coefficients of equation(s) of the level just below. At each level, a possible variation is modelled. That way, the dependencies in the data are taken into account, which is a first strength of this approach. By estimating and testing the different-level variances, it can be investigated if the overall intercept and slope vary between for instance participants, classes, studies, schools, et cetera. As a second strength, the variability can be explored by including characteristics of studies, schools, participants, and so on in the regression equations, and testing if these variables can partially explain the observed differences. Third, the model is very flexible and can be adapted to the unique investigated situation (Van den Noortgate & Onghena, 2003a, 2008). In general, multilevel models imply that identified characteristics can explain the observed differences, but only to some extent. They account for the uncertainty about possible moderators and combine both fixed-effects (e.g., Rosenthal, 1991) and random-effects (e.g., DerSimonian & Laird, 1986) models (Van den Noortgate & Onghena, 2003b, 2003c).

Most published SCE meta-analyses on interventions for CB among persons with DD apply nonregression approaches and calculate for instance the percentages of non-overlapping data (PND; e.g., Campbell, 2003, 2004; Didden, Korzilius, van Oorsouw, & Sturmey, 2006) or the improvement rate differences (IRD; Ganz et al., 2012; Parker, Vannest, & Brown, 2009) for all included cases and/or studies. Only some SCE meta-analyses on interventions for CB in persons with DD use multilevel models to analyse the data (e.g., Denis, Van den Noortgate, & Maes, 2011; Heyvaert et al., 2012).

To conduct the current multilevel meta-analysis of SCE data, the restricted maximum likelihood method was used through the procedure MIXED from SAS (Version 9.3; Littell,

Milliken, Stroup, Wolfinger, & Schabenberger, 2006; SAS Institute, 2011). We estimated and tested the overall intercept and the overall effect of psychosocial interventions for VCB, and estimated and tested the (co)variances over participants and studies. Subsequently, based on a priori considerations and guided by the data (i.e., significance tests, estimated moderator effects, fit indices: the Akaike Information Criterion (*AIC*), the corrected Akaike Information Criterion (*AICC*), and the Schwarz Bayesian Information criterion *(BIC)*), we built and tested several models containing various potential moderators of the treatment effect in order to construct the final three-level model with seven potential moderators (cf. Appendix D; Ferron et al., 2004). We retained the best model and studied in detail whether the intervention effect depended on the values of potential moderating variables (cf. Table 1 and Table 2; cf. Snijders & Bosker, 2012; Van den Noortgate & Onghena, 2008). Fifty-two studies reporting on 74 participants were included in the meta-analysis. Consequently, it was possible to apply a multilevel approach and to obtain quite precise parameter estimates, despite of the small amount of observations for some cases (cf. Van den Noortgate & Onghena, 2003a, 2007).

In addition, a sensitivity analysis and a publication bias analysis were conducted using the Metafor Package in R (Viechtbauer, 2010). More specifically, we checked for extreme outliers by generating a box and whisker diagram of the standardised random effects for the individual participants and applied the three-level models to the datasets without these outliers to investigate the robustness of the results. Finally, we drew a funnel plot (Rothstein, Sutton, & Borenstein, 2005) and tested for funnel plot asymmetry in order to assess the potential presence of publication bias (Egger, Smith, Schneider, & Minder, 1997).

3. Results

3.1. Three-level Model

3.1.1. Three-level model without moderators. First, we consider the results of the multilevel analysis without moderators, presented as Model 1 in Table 1 (cf. Van den Noortgate & Onghena, 2008). The estimated overall effect is –2.34. Accordingly, the level of VCB is on average 2.34 standard deviations lower during the treatment conditions, compared to the baseline conditions. According to a Wald test, this is a statistically significant decrease, $Z = –9.53$, $p < .0001$ (cf. research question 1). Further, we see that the treatment effects vary significantly over studies, $Z = 1.84$, $p = .0328$. The estimated variance over studies is 1.74 (cf. research question 2). The treatment effects also vary significantly over the included participants, $Z = 2.46$, $p = .0069$, with an estimated variance of 1.58 (cf. research question 3).

Table 1

Parameter Estimates and Standard Errors for the Multilevel Meta-analysis of SCEs

	Model 1	Model 2	Model 3
Regression coefficients			
Mean intervention effect	–2.34 (0.25)***		–1.88 (0.57)**
Moderator effect of			
Age		0.00 (0.03)	
Gender		–0.57 (0.52)	
VCB type		0.64 (0.25)**	0.62 (0.23)**
Intervention type		–0.80 (0.21)***	–0.78 (0.20)***
Primary treatment setting		0.26 (0.32)	
Publication year		0.03 (0.06)	
Study quality		0.42 (0.27)	
Variance of effect			
Between studies[a]	1.74 (0.94)*	2.07 (1.28)	2.25 (1.09)*
Between participants[b]	1.58 (0.64)**	1.87 (0.81)*	1.71 (0.67)**
Residual variance	1.00 (0.02)***	0.96 (0.02)***	0.96 (0.02)***

Notes. Means and variances of the intercept and covariances between studies and participants are not presented.
[a]Variance of effect between studies for Model 1 for the dataset without the two outliers: 0.70 (0.37)*. Variance of effect between studies for the dataset without the most outlying case: 0.43 (0.36).
[b]Variance of effect between participants for Model 1 for the dataset without the two outliers: 0.81 (0.30)**. Variance of effect between participants for the dataset without the most outlying case: 1.27(0.39)***.
*$p < .05$; **$p < .01$; ***$p < .001$.

3.1.2. Three-level models with moderators. Second, we examine which variables can explain the variability of the effect over studies and participants (cf. research question 4).

3.1.2.1. Model development. As mentioned in Section 2.2.1, we coded the included articles for several potential moderating variables. Due to missing data, we could not test some possible moderators (cf. Appendix C). As a result, we had 11 testable potential moderators: age, gender, VCB type, FA, intervention type, primary treatment setting, primary intervention agent, design, treatment datapoints, publication year, and study quality. Since the dataset contained the data of 74 participants, we aimed at building a three-level model with a maximum of seven potential moderators. Including a lot of possible moderators can lead to non-detection of significant effects (Lipsey, 2003). To select the remaining variables and to obtain the definitive Model 2 (cf. Table 1), four steps were taken, whereby a priori considerations were combined with data-guided decisions (cf. Ferron et al., 2004). The process of model development and the discussed results are displayed in detail in Appendix D. The names of the models refer to the number of potential moderating variables included.

We started with a three-level model containing the variables age, gender, publication year, and study quality (cf. Step I in Appendix D). Since these variables are included in multiple SCE meta-analyses on interventions for CB among individuals with DD (e.g., Campbell, 2003; Denis et al., 2011; Didden et al., 2006; Heyvaert et al., 2012, Heyvaert, Saenen, Maes, & Onghena, submitted), we considered them as "standard" meta-analytical variables. None of them seemed a significant moderator of the treatment effect. Subsequently, each of the other seven potential moderators was separately added to the model with the "standard" meta-analytical variables (cf. Step II in Appendix D). We kept the variables that were significant moderators in their separate five-variable models (i.e., VCB type and intervention type) and combined them with the "standard" meta-analytical variables in a six-variable model (cf. Step III in Appendix D). Also in this model, both VCB type and intervention type showed a statistically significant moderating effect. Next, we added every remaining variable separately to this six-variable model (cf. Step IV in Appendix D). None of them turned out to be an additional moderator. However, the seven-variable model containing primary treatment setting had better fit indices (*AIC*, *AICC*, and *BIC*) than the named six-variable and the other seven-variable models. Therefore, and to pay attention to the treatment setting (interventions conducted in classrooms have to face several unique challenges; cf. Machalicek et al., 2007), the seven-variable model including primary treatment setting was presented as the definitive Model 2 (cf. Table 1). During the model construction process, we built several models containing FA (cf. Appendix D). The sign of the estimated moderator effect of FA varied remarkable between these models. Because of data-analytical problems and a plausible multicollinearity, we excluded FA from the final model (cf. Lipsey, 2003).

3.1.2.2. Examining the developed model. Let us study Model 2, the three-level model with seven potential moderators (cf. Van den Noortgate & Onghena, 2008), presented in Table 1. In Model 2, both VCB type, $Z = 2.59$, $p = .0096$, and intervention type, $Z = -3.83$, $p = .0001$, significantly moderate the treatment effect. The variables age, $Z = 0.04$, $p = .9700$, gender, $Z = -1.09$, $p = .2762$, primary treatment setting, $Z = 0.80$, $p = .4258$, publication year, $Z = 0.47$, $p = .6384$, and study quality, $Z = 1.55$, $p = .1214$, do not have a significant moderating effect. Nevertheless, also this seven-variable model or the previously tested four-variable and five-variable models might include too many variables to identify moderating effects (cf. Lipsey, 2003). Hence, we added every potential moderator of Model 2 separately to the three-level model. These analyses bring the same conclusions as previously found. VCB type, $Z = 3.16$, $p = .0016$, and intervention type, $Z = -3.98$, $p < .0001$, significantly moderate the treatment effect. Further, no significant moderator effect is seen for age,

Z = 0.49, p = .6244, gender, Z = –0.81, p = .4180, primary treatment setting, Z = 0.72, p = .4733, publication year, Z = 1.19, p = .2357, and study quality, Z = 0.76, p = .4476.

Subsequently, only the significant moderators of the treatment effect in Model 2 were simultaneously examined in Model 3 (cf. Van den Noortgate & Onghena, 2008). Note that the same Model 3 would have been the result when we had retained the six-variable model (i.e., the three-level model without primary treatment setting; cf. Step III in Appendix D) as the definitive Model 2. In Model 3, both VCB type, Z = 2.75, p = .0060, and intervention type, Z = –3.89, p = .0001, remain to have significant moderating effects on the treatment effect. Accordingly, we conclude that there are differences in treatment effectiveness between the various types of VCB and between the different intervention types. A more detailed investigation of these moderating variables is needed.

3.1.2.3. Further examining VCB type and intervention type. To examine the meaning of the moderating effects of VCB type and intervention type, we studied the different VCB types and intervention types in two separate models, excluding other variables (respectively Model 4 and Model 5 in Table 2).

Looking at Model 4, we notice that the estimated overall intervention effect is –2.56, Z = –8.98, p < .0001. Compared to baseline conditions, the level of VCB is on average 2.56 standard deviations lower during treatment conditions. This is also the expected intervention effect for participants with stereotypical VCB (i.e., –2.56 + 0). Because stereotypical VCB was used as the base category against which the parameters of the other types were estimated and tested, its estimated moderator effect is equal to zero. The expected effect for participants with non-stereotypical VCB is only –2.05 (i.e., –2.56 + 0.51), which significantly differs from the effect for participants with stereotypical VCB, Z = 2.08, p = .0373. The expected intervention effect for participants with combined VCB is –2.43 (i.e., –2.56 + 0.13), which is not significantly different from the expected effects for participants with stereotypical and non-stereotypical VCB, respectively Z = 0.17, p = .8648; Z = –0.47, p = .6389.

Examining Model 5, we see that the average treatment effect is –1.94, Z = –5.87, p < .0001, which is also the expected effect of antecedent interventions (i.e., –1.94 + 0). Since antecedent intervention was used as the base category, its estimated moderator effect is equal to zero. The expected effect of consequence interventions is –2.27 (i.e., –1.94 + (–0.33)), which does not significantly differ from that of antecedent interventions, Z = –0.84, p = .3989. However, combined interventions, with an expected effect of –3.05 (i.e., –1.94 + (–1.11)), on average turn out to be significantly more effective than both single antecedent and consequence interventions, respectively Z = –2.69, p = .0072; Z = –2.41, p = .0159.

Table 2

Parameter Estimates and Standard Errors for the Multilevel Meta-analysis of SCEs: Further Examining VCB Type and Intervention Type

Model 4		Model 5	
Regression coefficients		Regression coefficients	
Mean intervention effect	–2.56 (0.29)***	Mean intervention effect	–1.94 (0.33)***
Moderator effect of		Moderator effect of	
Stereotypical VCB[a]	0.00 (-)	Antecedent[b]	0.00 (-)
Non-stereotypical VCB	0.51 (0.24)*	Consequence	–0.33 (0.39)
Combined VCB	0.13 (0.79)	Combined intervention	–1.11 (0.41)**
Variance of effect		Variance of effect	
Between studies	2.01 (0.99)*	Between studies	1.93 (1.01)*
Between participants	1.65 (0.66)**	Between participants	1.57 (0.64)**
Residual variance	0.99 (0.02)***	Residual variance	0.97 (0.02)***

Notes. Means and variances of the intercept and covariances between studies and participants are not presented.
[a]Stereotypical VCB was used as the base category against which the parameters of the other types of VCB were estimated and tested. In addition, to test the difference between non-stereotypical VCB and combined VCB, non-stereotypical VCB was used as the base category. Estimated moderator effect of combined VCB: –0.37 (0.79).
[b]Antecedent intervention was used as the base category against which the parameters of the other intervention types were estimated and tested. In addition, to test the difference between consequence and combined interventions, consequence intervention was used as the base category. Estimated moderator effect of combined intervention: –0.78 (0.32)*.
*$p < .05$; **$p < .01$; ***$p < .001$.

3.2. Sensitivity Analysis

Figure 1 presents the box and whisker diagram of the standardised random effects for the 74 participants. For two (from separate studies) of the three outlying dots, the effect deviates more than three standard deviations from zero (i.e., the average). Consequently, these cases can be seen as “extreme” outliers. Compared to the other cases, the most extreme outlier has only a limited number of observations (i.e., eight), which is indeed associated with outcomes that are further situated from the average effect (cf. Section 3.3 and Figure 2). To examine the influence of these outliers on the conclusions, we conducted a sensitivity analysis. We compared the results of the full dataset with the results of the dataset without the two outliers, and with the results of the dataset without the most extreme outlier.

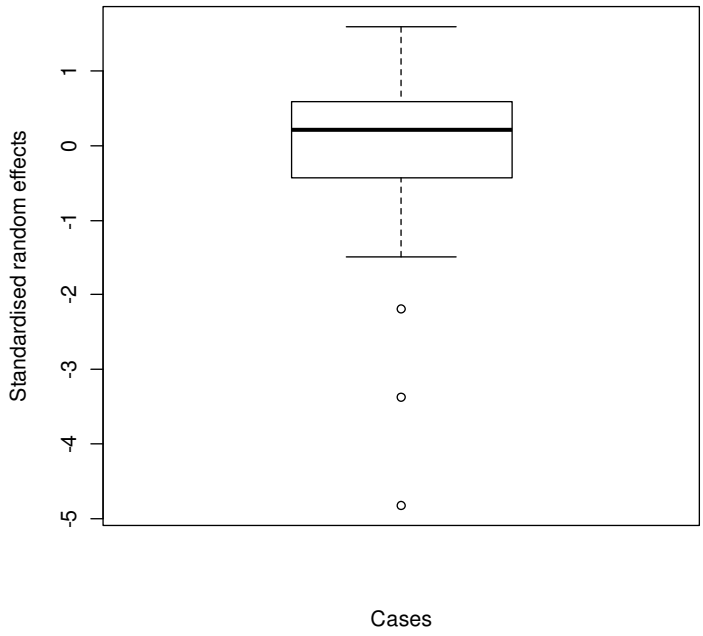

Figure 1. Box and whisker diagram of the standardised random effects.

To start, we applied Model 1 to the two new datasets (cf. Section 3.1.1). The overall intervention effect of –2.34 decreases to –2.11 for the dataset without the two outliers and to –2.13 for the dataset without the most extreme outlier. These three reductions are statistically significant according to a Wald test, with respectively $Z = –9.53$, $p < .0001$; $Z = –12.47$, $p < .0001$; $Z = –12.34$, $p < .0001$. Apparently, for each dataset, psychosocial interventions are on average highly effective in reducing VCB in individuals with autistic disorder. Subsequently, we consider the variability over studies, also enclosed in Model 1. We mentioned an estimated variance of 1.74 between studies, $Z = 1.84$, $p = .0328$, for the first dataset. By removing the two outliers, the estimated variance decreases more than half and is equal to 0.70. However, the effects still vary significantly over studies, $Z = 1.89$, $p = .0291$. After omitting only the most extreme case, the estimated variance is even reduced to 0.43, whereby treatment effects no longer significantly differ between studies, $Z = 1.19$, $p = .1167$. Further, we study the variability over participants. For the first dataset, we reported an estimated variance of 1.58, $Z = 2.46$, $p = .0069$. After omitting the two outliers, the estimated variance decreases almost by half and is equal to 0.81. Nevertheless, the effects still vary significantly over the included participants, $Z = 2.70$, $p = .0035$. By removing only the most extreme case, the estimated variance decreases to 1.27. Thereby, intervention effects still vary significantly over the included participants, $Z = 3.29$, $p = .0005$.

Subsequently, we applied Model 2 containing age, gender, VCB type, intervention type, primary treatment setting, publication year, and study quality (cf. Section 3.1.2.2) to the two new databases. In the original database VCB type, $Z = 2.59$, $p = .0096$, and intervention type, $Z = –3.83$, $p = .0001$, turned out to be significant moderators of the treatment effect. For the database without the two outliers, VCB type, $Z = 2.33$, $p = .0198$, and intervention type, $Z = –3.70$, $p = .0002$, remain to show a significant moderating effect. Moreover, also study quality has a statistically significant moderator effect, $Z = 2.19$, $p = .0285$. Interventions on average turn out to be less effective (i.e., have smaller intervention effects) for cases of studies with a higher total score on the SCED Scale. Other variables do not have a significant moderating effect. Similar results about VCB type, $Z = 2.50$, $p = .0125$, and intervention type, $Z = –3.99$, $p < .0001$, are seen for the database without the single case. Like in the original dataset, no other variables are found to have a statistically significant moderating effect.

Furthermore, we combined the significant moderators of Model 2 in a final three-level model, excluding other predictors (cf. Section 3.1.2.2). Thus, for each dataset, a Model 3 was built. For the original dataset, only VCB type and intervention type were included in Model 3. In this model, both variables turned out to be significant moderators, respectively $Z = 2.75$, $p = .0060$; $Z = –3.89$, $p = .0001$. For the dataset without the two outliers, we tested a Model 3 containing VCB type, intervention type, and study quality. Both VCB type, $Z = 2.53$, $p = .0115$, and intervention type, $Z = –4.24$, $p < .0001$, remain to show a significant moderating effect. Study quality is no longer a significant moderator, $Z = 1.55$, $p = .1215$. For the dataset without the most extreme outlier, we tested a three-level model combining VCB type and intervention type. Similarly to Model 2, both VCB type and intervention type significantly moderate the intervention effect in Model 3, respectively $Z = 2.61$, $p = .0090$; $Z = –4.19$, $p < .0001$. A further examination (i.e., applying Model 4 and Model 5 to the two new datasets; cf. Section 3.1.2.3) reveals that interventions for VCB including stereotypical VCB and combined interventions have still, on average, the largest intervention effects.

3.3. Publication Bias Analysis

Figure 2 displays a funnel plot in which the observed outcome of each participant (x-axis) is plotted against the corresponding standard error, which is a measure of the study size or precision (y-axis; Sterne & Egger, 2001). The vertical line corresponds to the estimated intervention effect based on the multilevel model without moderators (Model 1). Around the estimate, a pseudo confidence interval with a margin of 1.96 times the standard error value is represented (Viechtbauer, 2010).

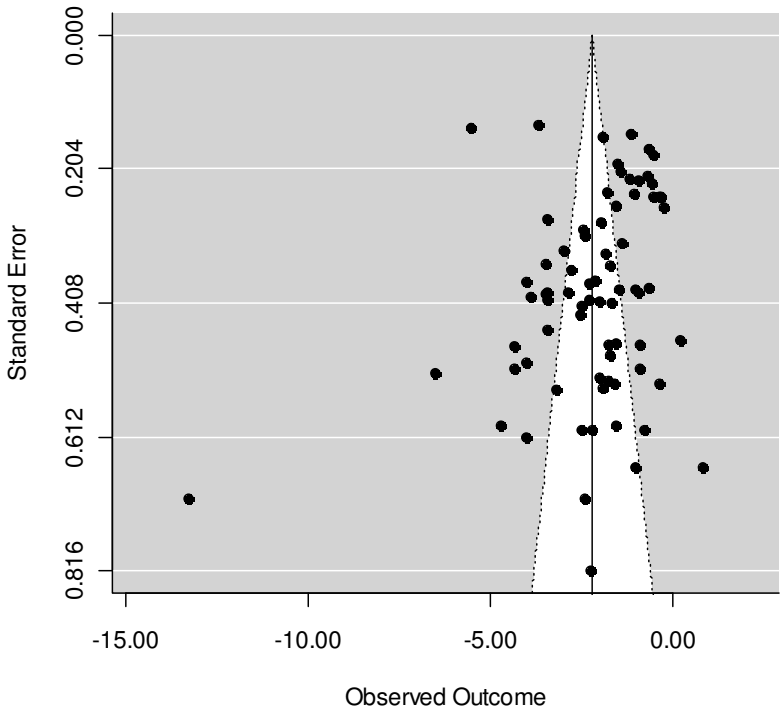

Figure 2. Funnel plot of the observed outcomes and the corresponding standard errors.

The logic behind such a plot is that the precision of the estimated effect increases when the study size (i.e., the number of observations) enlarges. The effects of larger studies (i.e., of cases with more observations), which are more powerful, are plotted more close to the overall treatment effect on top of the graph (since the y-axis is reversed). Smaller studies (i.e., cases with fewer observations) are plotted more broadly beneath. This produces an inverted funnel (Sterne et al., 2011). For several reasons, including publication bias, a funnel plot can lack symmetry and will be skewed (Egger, Smith et al., 1997).

By visual inspection, we notice only some skewness and limited evidence for publication bias. The plot is relatively symmetric with a small group of points on the right side on top of the plot and one extreme outlying point and one relatively outlying point on the left side of the plot. These two outlying cases on the left side were also identified in the box and whisker diagram (Figure 1). A regression test with standard error as predictor was carried out (cf. Viechtbauer, 2010). The test reveals that there is no statistically significant asymmetry in the funnel plot, t (72) = –0.64, p = .5213.

4. Discussion

Conducting a multilevel meta-analysis, we aggregated SCEs on the effectiveness of psychosocial interventions for VCB in individuals with autistic disorder. We investigated the

overall effect and the variability over studies and participants and examined the effect of potential moderators on the intervention effect. The overall intervention effect was large (–2.34) and statistically significant, $Z = –9.53$, $p < .0001$ (cf. research question 1). Hence, we conclude that the psychosocial interventions reported in the included SCEs were on average highly effective in reducing VCB in individuals with autistic disorder. These results confirm our hypothesis regarding the overall effect. The large average effect we found, is consistent with the effects reported in the SCE meta-analysis on intervention for CB in persons with autistic disorder of Campbell (2003). Also certain SCE meta-analyses and reviews on some of the interventions we studied, concluded that these interventions were effective in reducing CB (e.g., Chowdhury & Benson, 2011; Kokina & Kern, 2010; Mancil, 2006). In addition, also several SCE meta-analyses on psychosocial interventions for CB in individuals with DD found a large overall intervention effect (e.g., Didden et al., 2006; Harvey et al., 2009).

Nevertheless, we have to formulate three comments regarding the generalizability of our findings. First, it is important to keep in mind that the reported effect is an average effect. Moreover, the estimated intervention effect varied significantly over the included studies (the estimated variance is 1.74), $Z = 1.84$, $p = .0328$, and participants (the estimated variance is 1.58), $Z = 2.46$, $p = .0069$ (cf. research questions 2 and 3). Nonetheless, these variances are quite small. To obtain a more detailed insight into the results, we calculated confidence intervals. Based on a normal distribution, we expect that 95% of the studies have outcomes between –4.93 (i.e., $–2.34 - 1.96 * \sqrt{1.74}$) and 0.25 (i.e., $-2.34 + 1.96 * \sqrt{1.74}$). Negative values mean that there is a decrease in VCB due to the psychosocial interventions, positive values indicate an increase in VCB. Equally, for a typical study, 95% of the participants are estimated to score within the range –4.80 (i.e., $-2.34 - 1.96 * \sqrt{1.58}$) till 0.12 (i.e., $-2.34 + 1.96 * \sqrt{1.58}$). Thus, though the variability over studies and participants was statistically significant, psychosocial interventions are expected to have the intended effect for the majority of the studies and participants. Moreover, after removing the outliers, the variances were even smaller. In addition, the variability over studies was no longer significant for the dataset without the most extreme outlier (cf. Section 3.2 and Table 1).

A second concern relates to publication bias. Since SCEs reporting positive treatment outcomes (i.e., a reduction in CB) are more likely to be published than SCEs reporting negative (i.e., an increase in CB) or absent (i.e., nor an increase, nor a decrease) treatment results, the findings of SCE meta-analyses can be biased (Egger, Smith et al., 1997; Thornton & Lee, 2000). The regression test confirmed the visual analysis and revealed that there was

no statistically significant asymmetry in the funnel plot (cf. Figure 2 and Section 3.3). We conclude that the present study does not suffer much from publication bias effects, although publication bias cannot be excluded absolutely (Sterne et al., 2011). Additionally, since we applied an English-language inclusion criterion, it is possible that the present results differ a little from the results of all the published research in this area. Optimistic outcomes are more often published in English-language journals (Egger, Zellweger-Zähner et al., 1997; Rothstein et al., 2005).

Third, it is important to note that the great majority of the included participants were children and adolescents (the average age of the included participants was 10.0 year). Therefore, the conclusions of this study can best be applied to the population of children and adolescents. Nevertheless, a moderating effect of age was not found. However, this can be the result of the limited variability in age between the studied participants (cf. Van den Noortgate & Onghena, 2007).

Finally, we have to make an additional comment regarding the systematic search process. Applying the search string: (*autism* OR *autistic*) AND (*vocal**), only 20 articles (21 articles were retained by the first author, one article was excluded afterwards) were retained by screening the four databases. These articles concerned for instance *vocal stereotypy*, *vocali(s)(z)ations of fear*, and *inappropriate vocali(s)(z)ations* and reported these terms in the title, abstract, or among the keywords. Subsequently, 32 (33 articles were retained by the first author, one article was excluded) additional articles were found during the manual search of 23 journals. Reflecting on the final database, this large quantity of manually found articles has several reasons. First, several forms of VCB are verbal. Consequently, articles use terms including *verbal* (e.g., *aberrant verbal behavio(u)r*, *interrupting verbali(s)(z)ations*) or *speech* (e.g., *irrelevant speech, perseverative speech*). Second, some studies use only more specific terms of VCB (e.g., *coprolalia*, *high-pitched voice*, *screaming*) in the title, abstract or keywords. Third, and on the contrary, various articles use a more general terminology (e.g., *disruptive/problem/inappropriate behavio(u)r*), to describe the subject of the study. Fourth, some reports of SCEs about interventions for VCB only mention the used interventions (e.g., *differential reinforcement*) in the title, abstract or keywords. We conclude that it could be useful to apply more specific search strings (e.g., *echolalia*, *scream**, *persever**) and combinations with *verbal* and *speech* to search for SCEs on interventions for VCB. Nevertheless, also a hand search of journals and attention for treatment terms and more general CB terms are needed to find as much as possible SCEs on the topic.

Despite its limitations, the present study has several strengths. The first advantage is formed by the study design, a meta-analysis of SCEs. By combining the results of several SCEs, this meta-analysis offers more reliable conclusions about the overall effect than individual SCEs and enlarges the power in testing this effect. Furthermore, besides information about the overall effect, information about moderating variables is obtained (cf. Van den Noortgate & Onghena, 2003a, 2003b). These results were achieved systematically and objectively, aspects that are not guaranteed in traditional narrative reviews, which can also contain biased results (Cooper, 2010; Green, Johnson, & Adams, 2006). In addition, a quantitative aggregation of SCEs offers information that neither separate SCEs, nor group studies can give: information about the overall effect as well as about specific cases (Van den Noortgate & Onghena, 2003a, 2003b).

The extensive moderator analyses can be considered as a second strength of the current study. Based on previous meta-analyses of SCEs on interventions for CB among individuals with autism and/or other DD, we created a list of potential moderating variables at the level of the individual, at the level of the intervention and the intervention context, and at the level of the study. We coded the included SCEs for these variables and calculated descriptive statistics (cf. Appendix C). Subsequently, we built several three-level models and developed the definitive Model 2 (cf. Appendix D) in order to investigate the effect of the remaining potential moderators (cf. Table 1 and Table 2; cf. research question 4). We hypothesised that all potential moderators in Appendix C could have a moderating effect, but that especially VCB type, intervention type, pretreatment FA, and publication year would moderate the intervention effect. By building and testing several models, we created the opportunity to examine multiple potential moderators, despite of the relatively "small" dataset. In addition, building and testing several models offered the opportunity to pay attention to VCB type, intervention type, pretreatment FA and the "standard" meta-analytical variables, without leaving out of account other relevant potential moderators (e.g., primary treatment setting, design).

Inconsistent with our expectations, the present study could not affirm that a pretreatment FA is associated with better treatment effects for VCB in individuals with autistic disorder. In contrast, previous reviews concluded that a pretreatment FA resulted in better outcomes for CB in individuals with autistic disorder (Campbell, 2003) and/or other DD (Didden et al., 2006; Harvey et al., 2009). In addition, we did not find a significant moderator effect of primary treatment setting, primary intervention agent, and study design. These variables had a significant moderating effect on the intervention effect in the SCE

meta-analysis focused on individuals with autism of Ma (2009). However, it is important to note that Ma (2009) included also data for positive behaviour. Furthermore, like in the SCE meta-analyses on autism of Campbell (2003) and Ma (2009), nor age or gender revealed to be significant moderators of the treatment effect in the current study. Even though the moderator effect of study quality was only significant in Model 2 for the dataset without the two outliers (cf. Section 3.2), we found that interventions were on average less effective (i.e., had smaller intervention effects) for cases of studies with a higher total score on the SCED Scale (cf. Table 1). In contrast, some quality aspects (e.g., reliability of observation, generalisation) were associated with better outcomes in the studies of Campbell (2003) and Didden et al. (2006). Finally, we did not find a significant moderating effect of publication year and treatment datapoints.

Although several variables did not have a significant moderating effect on the treatment effect, two variables were found to have a significant moderator effect: VCB type and intervention type (cf. Table 1 and Table 2). That way, the present study confirmed our hypotheses that especially VCB type and intervention type would moderate the intervention effect. Even after omitting the extreme outliers, both VCB type and intervention type still had a significant moderating effect (cf. Section 3.2). Unlike our meta-analysis, Campbell (2003) did not find significant moderating effects of CB type and intervention type. Nonetheless, we studied a specific type of CB in individuals with autistic disorder and made somewhat different categories. Since our categories were on the one hand broader (e.g., we had only three intervention types), and on the other hand more specific (e.g., we made a separate category for stereotypical VCB) than the categories of Ma (2009), more univocal conclusions about these moderators could be formulated.

Interventions aimed at decreasing stereotypical VCB (the expected intervention effect is –2.56) were on average significantly more effective than interventions for non-stereotypical VCB (the expected effect is only –2.05). The effect of interventions used to reduce combined VCB (the expected effect is –2.43) did not significantly differ from the effects of interventions for stereotypical and non-stereotypical VCB. Also in the SCE meta-analysis about behavioural interventions for CB in individuals with DD including also children with autism of Harvey et al. (2009), better treatment results were found for interventions used to reduce stereotypy, compared to interventions for other CBs such as disruptive behaviours. In addition, the SCE meta-analysis on behavioural interventions for persons with mild ID including also individuals with ASD by Didden et al. (2006) found relatively high mean effect sizes for stereotypical behaviour.

Finally, interventions including both preventing VCB as well as reacting on the presence or absence of VCB (the expected effect is –3.05) on average turned out to be significantly more effective than single antecedent (the expected effect is –1.94) or consequence interventions (the expected effect is –2.27). The effects of antecedent and consequent interventions did not significantly differ from one another. This research finding can be very valuable for practitioners who intend to reduce VCB in persons with autistic disorder. Namely, based on these results, we expect that interventions such as social stories, (video) modelling, or noncontingent reinforcement combined with for example contingent praise or remarks will be more effective in reducing VCB in persons with autistic disorder than single modelling, cue card, weighted vest, snoezelen, or prompting interventions, and than solely differential reinforcement, punishment, self-management, or functional communication training.

Before we discuss the next strength, we have to make a supplementary comment concerning moderating variables. As mentioned in Section 3.1.2.1 and in Appendix C, we could not examine the potential moderating effect of several variables. First of all, certain participant characteristics were only reported in some studies (i.e., the use of medication, ID level, and additional language problem) and were therefore not incorporated in the remaining analyses. Second, also for the variables duration of treatment and frequency of treatment, a lot of missing values were scored. Some studies did not mention the duration and/or the frequency of the intervention. Other studies used different ways of reporting these aspects (e.g., minimum and maximum number of sessions/day, number of days/week, number of school years). Unless we searched for a straightforward way of coding (i.e., calculating the median and using information from the graph), it was for several studies simply not feasible to deduce unequivocal and comparative information about these variables. Thus, due to missing values and problems with similarities between studies, it was not possible to present a complete overview of participant and intervention characteristics. As a result, the potential moderator effect of some variables could not be examined. For that reason, and line with Campbell (2003), Denis et al. (2011), and Harvey et al. (2009), we would encourage researchers to report more characteristics of the participants in SCEs and to incorporate more clear and comparable information about the duration and the frequency of the studied interventions. Accordingly, guidelines for reporting these aspects would be useful.

As a third strength, we mention that the interventions still revealed to be on average highly effective after removing the most extreme outliers. Moreover, the conclusions about the variances over studies and participants did not fundamentally change after removing the

outliers. In addition, VCB type and intervention type still significantly moderated the intervention effect when these outliers were removed (cf. Section 3.2). We conclude that the outcomes of the present meta-analysis are quite robust.

Fourth, by including only persons with autistic disorder, we selected a quite homogenous population and made clinical applications of our results more straightforward. As a result, the present results about psychosocial interventions for VCB in individuals with autistic disorder thus cannot be generalised to other individuals (e.g., individuals with only ID), other types of CB, or to pharmacological interventions for VCB.

We summarise that the overall effect of psychosocial interventions for VCB in individuals with autistic disorder was large and statistically significant. Although the variability was relatively small, the results varied significantly over the included studies and participants. Both VCB type and intervention type revealed to be statistically significant moderators of the intervention effect, with, on average, the largest intervention effects for interventions used to reduce VCB including stereotypical VCB and for interventions containing both antecedent and consequence components. Age, gender, primary treatment setting, publication year, and study quality did not significantly moderate the treatment effect. These results were quite robust: removing the most extreme outliers did not substantially change the results.

The present meta-analysis is the first review that focuses on both vocal stereotypy and other VCBs in individuals with autistic disorder. Over the last years, the number of SCE studies on VCB among individuals with autistic disorder is increasing remarkably (cf. Lanovaz & Sladeczek, 2011). Updates of the present research are thus needed to contribute to cumulative evidence and to well-founded interventions for VCB in individuals with autistic disorder. Nevertheless, based on the present SCE meta-analysis, we can be optimistic about the effectiveness of psychosocial interventions for VCB in individuals with autistic disorder.

Author Affiliation

Lien Vanderkerken is a master student in Educational Sciences, KU Leuven (Belgium).

Mieke Heyvaert is a postdoctoral fellow of the Research Foundation – Flanders (Belgium).

Bea Maes is a full professor at the Faculty of Psychology and Educational Sciences, KU Leuven (Belgium).

Patrick Onghena is a full professor at the Faculty of Psychology and Educational Sciences, KU Leuven (Belgium).

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Athens, E. S., Vollmer, T. R., Sloman, K. N., & Pipkin, C. St. P. (2008). An analysis of vocal stereotypy and therapist fading. *Journal of Applied Behavior Analysis, 41,* 291–297. doi:10.1901/jaba.2008.41-291

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Allyn & Bacon.

Buschbacher, P. W., & Fox, L. (2003). Understanding and intervening with the challenging behavior of young children with autism spectrum disorder. *Language, Speech, and Hearing Services in Schools, 34,* 217–227. doi:10.1044/0161-1461(2003/018)

Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24,* 120–138. doi:10.1016/S0891-4222(03)00014-3

Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification, 28,* 234–246. doi:10.1177/0145445503259264

Chowdhury, M., & Benson, B. A. (2011). Use of differential reinforcement to reduce behavior problems in adults with intellectual disabilities: A methodological review. *Research in Developmental Disabilities, 32,* 383–394. doi:10.1016/j.ridd.2010.11.015

Cohen, I. L., Yoo, H. J., Goodwin, M. S., & Moskowitz, L. (2011). Assessing challenging behaviors in autism spectrum disorders: Prevalence, rating scales, and autonomic indicators. In J. L. Matson & P. Sturmey (Vol. Eds.), *International handbook of autism and pervasive developmental disorders* (pp. 247–270). New York, NY: Springer.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders, 2,* 469–479. doi:10.1016/j.rasd.2007.09.006

Denis, J., Van den Noortgate, W., & Maes, B. (2011). Self-injurious behavior in people with profound intellectual disabilities: A meta-analysis of single-case studies. *Research in Developmental Disabilities, 32,* 911–923. doi:10.1016/j.ridd.2011.01.014

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188.

Didden, R., Korzilius, H., van Oorsouw, W., & Sturmey, P. (2006). Behavioral treatment of challenging behaviors in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation, 111,* 290–298. doi:10.1352/0895-8017(2006)111[290:BTOCBI]2.0.CO;2

Došen, A., Gardner, W. I., Griffiths, D. M., King, R., & Lapointe, A. (2007). *Practice guidelines and principles: Assessment, diagnosis, treatment, and related support services for persons with intellectual disabilities and problem behaviour*. Centre of Consultation and Expertise, Gouda, The Netherlands.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634.

Egger, M., Zellweger-Zähner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomised controlled trials published in English and German. *Lancet, 350,* 326–329. doi:10.1016/S0140-6736(97)02419-7

Emerson, E. (2001). *Challenging behaviour: Analysis and intervention in people with severe intellectual disabilities* (2nd ed.). Cambridge: Cambridge University Press.

Ferron, J., Hess, M. R., Hogarty, K. Y., Dedrick, R. F., Kromrey, J. D., Lang, T. R., & Niles, J. (2004, April). *Hierarchical linear modeling: A review of methodological issues and applications.* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Franco, J. H., Lang, R. L., O'Reilly, M. F., Chan, J. M., Sigafoos, J., & Rispoli, M. (2009). Functional analysis and treatment of inappropriate vocalizations using a speech-generating device for a child with autism. *Focus on Autism and Other Developmental Disabilities, 24,* 146–155. doi:10.1177/1088357609338380

Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., ... Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6,* 578–588. doi:10.1016/j.rasd.2011.09.007

Galiatsatos, G. T., & Graff, R. B. (2003). Combining descriptive and functional analyses to assess and treat screaming. *Behavioral Interventions, 18,* 123–138. doi:10.1002/bin.133

Ganz, J. B., Earles-Vollrath, T. L., Heath, A. K., Parker, R. I., Rispoli, M. J., & Duran, J. B. (2012). A meta-analysis of single case research studies on aided augmentative and alternative communication systems with individuals with autism spectrum disorders.

Journal of Autism and Developmental Disorders, 42, 60–74, doi:10.1007/s10803-011-1212-2

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8. doi:10.3102/0013189X005010003

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5,* 101–117. doi:10.1016/S0899-3467(07)60142-6

Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trail: Gold standard, or merely standard? *Perspectives in biology and medicine, 48,* 516–534. doi:10.1353/pbm.2005.0092

Haahr, M. (1998). RANDOM.ORG. Retrieved from http://www.random.org/

Harvey, S. T., Boer, D., Meyer, L. H., & Evans, I. M. (2009). Updating a meta-analysis of intervention research with challenging behaviour: Treatment validity and standards of practice. *Journal of Intellectual and Developmental Disability*, *34,* 67–80. doi:10.1080.13668250802690922

Hastings, R. P. (2002). Parental stress and behaviour problems of children with developmental disability. *Journal of Intellectual and Developmental Disability, 27,* 149–160. doi:10.1080/1366825021000008657

Hastings, R. P. (2005). Staff in special education settings and behaviour problems: Towards a framework for research and practice. *Educational Psychology, 25,* 207–221. doi:10.1080/0144341042000301166

Healy, O., & Leader, G. (2011). Assessment of rituals and stereotypy. In J. L. Matson & P. Sturmey (Vol. Eds.), *International handbook of autism and pervasive developmental disorders* (pp. 233–246). New York, NY: Springer.

Helverschou, S. B., Bakken, T. L., & Martinsen, H. (2011). Psychiatric disorders in people with autism spectrum disorders: Phenomenology and recognition. In J. L. Matson & P. Sturmey (Vol. Eds.), *International handbook of autism and pervasive developmental disorders* (pp. 53–74). New York, NY: Springer.

Hetzroni, O. E., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of Autism and Developmental Disorders, 34,* 95–113. doi:10.1023/B:JADD.0000022602.40506.bf

Heyvaert, M., Maes, B., Van den Noortgate, W., Kuppens, S., & Onghena, P. (2012). A multilevel meta-analysis of single-case and small-n research on interventions for

reducing challenging behavior in persons with intellectual disabilities. *Research in Developmental Disabilities, 33,* 766–780. doi:10.1016/j.ridd.2011.10.010

Heyvaert, M., Saenen, L., Maes, B., & Onghena, P. (submitted). Systematic review of restraint interventions for challenging behaviour among persons with intellectual disabilities: Focus on effectiveness in single-case experiments. Manuscript submitted to the *Journal of Applied Research in Intellectual Disabilities*.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0) [updated March 2011, Electronic Version]. Retrieved from www.cochrane-handbook.org

Hodgetts, S., Magill-Evans, J., & Misiaszek, J. E. (2011). Weighted vests, stereotyped behaviors and arousal in children with autism. *Journal of Autism and Developmental Disorders*, *41,* 805–814. doi:10.1007/s10803-010-1104-x

Holden, B., & Gitlesen, J. P. (2009). The overlap between psychiatric symptoms and challenging behaviour: A preliminary study. *Research in Developmental Disabilities, 30*, 210–218. doi:10.1016/j.ridd.2008.03.003

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Howlin, P. (1998). Practitioner review: Psychological and educational treatments for autism. *Journal of Child Psychology and Psychiatry, 39,* 307–322. doi:10.1111/1469-7610.00327

Hutchins, T. L., & Prelock, P. A. (n.d.). Using communication to reduce challenging behaviors in individuals with autism spectrum disorders. Retrieved from http://www.vermontfamilynetwork.org/wp-content/uploads/2013/04/psychiatric-article-submitted-1_12_13.pdf

Iwakabe, S., & Gazzola, N. (2009). From single-case studies to practice-based knowledge: Aggregating and synthesizing case studies. *Psychotherapy Research, 19,* 601–611. doi:10.1080/10503300802688494

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Keeling, K., Smith Myles, B., Gagnon, E., & Simpson, R. L. (2003). Using the power card strategy to teach sportsmanship skills to a child with autism. *Focus on Autism and Other Developmental Disabilities, 18,* 105–111. doi:10.1177/108835760301800204

Kokina, A., & Kern, L. (2010). Social Story™ interventions for students with autism spectrum disorders: A meta-analysis. *Journal of Autism and Developmental Disorders, 40,* 812–826. doi:10.1007/s10803-009-0931-0

Langen, M., Durston S., Kas, M. J., van Engeland, H., & Staal W. G. (2011). The neurobiology of repetitive behavior: … and men. *Neuroscience and Biobehavioral Reviews. 35,* 356–365. doi:10.1016/j.neubiorev.2010.02.005

Lanovaz, M. J., & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification, 36,* 146–164. doi:10.1177/0145445511427192

Lequia, J., Machalicek, W., & Rispoli, M. J. (2012). Effects of activity schedules on challenging behavior exhibited in children with autism spectrum disorders: A systematic review. *Research in Autism Spectrum Disorders, 6,* 480–492. doi:10.1016/j.rasd.2011.07.008

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science, 587,* 69–81. doi:10.1177/0002716202250791

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS® for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.

Liu-Gitz, L., & Banda, D. R. (2010). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. *Behavioral Interventions, 25,* 77–87. doi:10.1002/bin.297

Ma, H.-H. (2009). The effectiveness of intervention on the behavior of individuals with autism: A meta-analysis using percentage of data points exceeding the median of baseline phase (PEM). *Behavior Modification, 33,* 339–359. doi:10.1177/0145445509333173

MacDonald, R., Green, G., Mansfield, R., Geckeler, A., Gardenier, N., Anderson, J., … Sanchez, J. (2007). Stereotypy in young children with autism and typically developing children. *Research in Developmental Disabilities, 28,* 266–277. doi:10.1016/j.ridd.2006.01.004

Machalicek, W., O'Reilly, M. F., Beretvas, N., Sigafoos, J., & Lancioni, G. E. (2007). A review of interventions to reduce challenging behavior in school settings for students with autism spectrum disorders. *Research in Autism Spectrum Disorders, 1,* 229–246. doi:10.1016/j.rasd.2006.10.005

Mancil, R. C. (2006). Functional communication training: A review of the literature related to children with autism. *Education and Training in Developmental Disabilities, 41,* 213–224.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30,* 1107–1114. doi:10.1016/j.ridd.2009.06.003

Matson, J. L., Turygin, N. C., Beighley, J., & Matson, M. L. (2012). Status of single-case research designs for evidence-based practice. *Research in Autism Spectrum Disorders 6,* 931–938. doi:10.1016/j.rasd.2011.12.008

Miguel, C. F., Clark, K., Tereshko, L., & Ahearn, W. H. (2009). The effects of response interruption and redirection and sertraline on vocal stereotypy. *Journal of Applied Behavior Analysis, 42,* 883–888. doi:10.1901/jaba.2009.42-883

Norris, C., & Dattilo, J. (1999). Evaluating effects of a social story intervention on a young girl with autism. *Focus on Autism and Other Developmental Disabilities, 14,* 180–186. doi:10.1177/108835769901400307

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75,* 135–150.

Rapp, J. T. (2007). Further evaluation of methods to identify matched stimulation. *Journal of Applied Behavior Analysis, 40,* 73–88. doi:10.1901/jaba.2007.142-05

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis. Prevention, assessment, and adjustments.* Chichester: John Wiley & Sons.

SAS Institute. (2011). *SAS/STAT user's guide* (Version 9.3) [Electronic Version]. Retrieved from http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#titlepage.htm

Schalock, R. L. (2004). The concept of quality of life: What we know and do not know. *Journal of Intellectual Disability Research, 48,* 203–216. doi:10.1111/j.1365-2788.2003.00558.x

Shadish, W. R., Brasil, I. C. C., Illingworth, D. A. I., White, K. D., Galindo, R., Nagler, E. D., & Rindskopf, D. M. (2009). Using UnGraph to extract data from image files:

Verification of reliability and validity. *Behavior Research Methods, 41,* 177–183. doi:10.3758/BRM.41.1.177

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,* 228–237. doi:10.1177/10983007040060040401

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54,* 1046–1055. doi:10.1016/S0895-4356(01)00377-8

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., … Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, *342*. doi:10.1136/bmj.d4002

Tarbox, J., Wilke, A. E., Najdowski, A. C., Findel-Pyles, R. S., Balasanyan, S., Caveney, A. C., … Tia, B. (2009). Comparing indirect, descriptive, and experimental functional assessments of challenging behavior in children with autism. *Journal of Developmental and Physical Disabilities, 21,* 493–514. doi:10.1007/s10882-009-9154-8

Tate, R. L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation, 18,* 385–401. doi:10.1080/09602010802009201

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53,* 207-216. doi:10.1016/S0895-4356(99)00161-4

Tomanik, S., Harris, G. E., & Hawkins, J. (2004). The relationship between behaviours exhibited by children with autism and maternal stress. *Journal of Intellectual and Developmental Disability, 29,* 16–26. doi:10.1080/13668250410001662892

UnGraph (Version 5.0) [Computer software]. Available from http://www.biosoft.com/w/ungraph.htm

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18,* 325–346. doi: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, & Computers*, *35,* 1–10.

Van den Noortgate, W., & Onghena, P. (2003c). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*, 765–790. doi:10.1177/0013164403251027

Van den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today, 8,* 196–209.

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48.

Vollmer, T. R. (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities, 15,* 187–207. doi:10.1016/0891-4222(94)90011-6

World Health Organization. (2013). *ICD-10* (Version 2010). Retrieved from http://apps.who.int/classifications/icd10/browse/2010/en

Zhan, S., & Ottenbacher, K. J. (2001). Single subject research designs for disability research. *Disability and Rehabilitation*, *23,* 1–8. doi:10.1080/09638280150211202

Appendix A

Table A.1

Overview of Journals and Corresponding Included Articles

Journal	*N*	Journal	*N*
Autism	0	*Journal of Abnormal Child Psychology**	0
Autism Research	0	*Journal of Applied Behavior Analysis**	21
*Behavior Modification**	3	*Journal of Applied Research in Intellectual Disabilities*[d]	0
*Behavior Therapy**	0	*Journal of Autism and Developmental Disorders**	6
*Behavioral Disorders**	0	*Journal of Behavior Therapy and Experimental Psychiatry**	0
*Behavioral Interventions**[, a]	7	*Journal of Clinical Child and Adolescent Psychology**	0
*Behaviour Research and Therapy**	0	*Journal of Consulting and Clinical Psychology**	0
*Child & Family Behavior Therapy**[, b]	1	*Journal of Experimental Child Psychology**	0
Clinical Case Studies	1	*Journal of Positive Behavior Interventions*	1
*Education and Training in Autism and Developmental Disabilities**[, c]	2	*Research in Autism Spectrum Disorders*	1
Focus on Autism and Other Developmental Disabilities	7	*Research in Developmental Disabilities**	2
*Intellectual and Developmental Disabilities**	0		

Notes. N = Number of articles that were included in the meta-analysis.
*Journals indicated with an asterisk were also integrated in the search process of Campbell (2003).
[a, d]For *Behavioral Interventions* and *Journal of Applied Research in Intellectual Disabilities,* volumes published since 1997 were manually searched. Previous volumes were not available.
[b]For *Child & Family Behavior Therapy*, volumes published in 1990 and volumes published since 1997 were manually searched. Other volumes were not available.
[c]Volumes 1 and 4 of 1997 were not available for *Education and Training in Autism and Developmental Disabilities.*

Appendix B

Reference List of the Included Articles

Ahearn, W. H., Clark, K. M., DeBar, R., & Florentino, C. (2005). On the role of preference in response competition. *Journal of Applied Behavior Analysis, 38,* 247–250. doi:10.1901/jaba.2005.36-04

Ahearn, W. H., Clark, K. M., MacDonald, R. P. F., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 40,* 263–275. doi:10.1901/jaba.2007.30-06

Ahrens, E. N., Lerman, D. C., Kodak, T., Worsdell, A. S., & Keegan, C. (2011). Further evaluation of response interruption and redirection as treatment for stereotypy. *Journal of Applied Behavior Analysis, 44,* 95–108. doi:10.1901/jaba.2011.44-95

Anderson, J., & Le, D. D. (2011). Abatement of intractable vocal stereotypy using an overcorrection procedure. *Behavioral Interventions, 26,* 134–146. doi:10.1002/bin.326

Arntzen, E., Tønnessen, I. R., & Brouwer, G. (2006). Reducing aberrant verbal behavior by building a repertoire of rational verbal behavior. *Behavioral Interventions, 21,* 177–193. doi:10.1002/bin.220

Athens, E. S., Vollmer, T. R., Sloman, K. N., & Pipkin, C. St. P. (2008). An analysis of vocal stereotypy and therapist fading. *Journal of Applied Behavior Analysis, 41,* 291–297. doi:10.1901/jaba.2008.41-291

Cassella, M. D., Sidener, T. M., Sidener, D. W., & Progar, P. R. (2011). Response interruption and redirection for vocal stereotypy in children with autism: A systematic replication. *Journal of Applied Behavior Analysis, 44,* 169–173. doi:10.1901/jaba.2011.44-169

Chan, J. M., & O'Reilly, M. F. (2008). A social stories™ intervention package for students with autism in inclusive classroom settings. *Journal of Applied Behavior Analysis, 41,* 405–409. doi:10.1901/jaba.2008.41-405

Coleman, C. L., & Holmes, P. A. (1998). The use of noncontingent escape to reduce disruptive behaviors in children with speech delays. *Journal of Applied Behavior analysis, 31,* 687–690. doi:10.1901/jaba.1998.31-687

Crozier, S., & Tincani, M. J. (2005). Using a modified social story to decrease disruptive behavior of a child with autism. *Focus on Autism and Other Developmental Disabilities, 20,* 150–157. doi:10.1177/10883576050200030301

Donaldson, J. M., & Vollmer, T. R. (2011). An evaluation and comparison of time-out procedures with and without release contingencies. *Journal of Applied Behavior Analysis, 44,* 693–705. doi:10.1901/jaba.2011.44-693

Earles, T. L., & Smith Myles, B. (1994). Using behavioral interventions to decrease coprolalia in a student with Tourette's syndrome and autism: A case study. *Focus on Autism and Other Developmental Disabilities, 8,* 1–10. doi:10.1177/108835769400800601

Falcomata, T. S., Roane, H. S., Hovanetz, A. N., Kettering, T. L., & Keeney, K. M. (2004). An evaluation of response cost in the treatment of inappropriate vocalizations maintained by automatic reinforcement. *Journal of Applied Behavior Analysis, 37,* 83–87. doi:10.1901/jaba.2004.37-83

Franco, J. H., Lang, R. L., O'Reilly, M. F., Chan, J. M., Sigafoos, J., & Rispoli, M. (2009). Functional analysis and treatment of inappropriate vocalizations using a speech-generating device for a child with autism. *Focus on Autism and Other Developmental Disabilities, 24,* 146–155. doi:10.1177/1088357609338380

Galiatsatos, G. T., & Graff, R. B. (2003). Combining descriptive and functional analyses to assess and treat screaming. *Behavioral Interventions, 18,* 123–138. doi:10.1002/bin.133

Ganz, J. B., Kaylor, M., Bourgeois, B., & Hadden, K. (2008). The impact of social scripts and visual cues on verbal communication in three children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 23,* 79–94. doi:10.1177/1088357607311447

Graetz, J. E., Mastropieri, M. A., & Scruggs, T. E. (2009). Decreasing inappropriate behaviors for adolescents with autism spectrum disorders using modified social stories. *Education and Training in Developmental Disabilities, 44,* 91–104.

Graff, R. B., Lineman, G. T., Libby, M. E., & Ahearn, W. H. (1999). Functional analysis and treatment of screaming in a young girl with severe disabilities. *Behavioral Interventions, 14,* 233–239. doi:10.1002/(SICI)1099-078X(199910/12)14:4<233::AID-BIN39>3.0.CO;2-P

Groskreutz, M. P., Groskreutz, N. C., & Higbee, T. S. (2011). Response competition and stimulus preference in the treatment of automatically reinforced behavior: A comparison. *Journal of Applied Behavior Analysis, 44,* 211–215. doi:10.1901/jaba.2011.44-211

Haley, J. L., Heick, P. F., & Luiselli, J. K. (2010). Use of an antecedent intervention to decrease vocal stereotypy of a student with autism in the general education classroom. *Child and Family Behavior Therapy, 32,* 311–321. doi:10.1080/07317107.2010.515527

Hetzroni, O. E., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of Autism and Developmental Disorders, 34,* 95–113. doi:10.1023/B:JADD.0000022602.40506.bf

Hodgetts, S., Magill-Evans, J., & Misiaszek, J. E. (2011). Weighted vests, stereotyped behaviors and arousal in children with autism. *Journal of Autism and Developmental Disorders, 41,* 805–814. doi:10.1007/s10803-010-1104-x

Kaplan, H., Clopton, M., Kaplan, M., Messbauer, L., & McPherson, K. (2006). Snoezelen multi-sensory environments: Task engagement and generalization. *Research in Developmental Disabilities, 27,* 443–455. doi:10.1016/j.ridd.2005.05.007

Keeling, K., Smith Myles, B., Gagnon, E., & Simpson, R. L. (2003). Using the power card strategy to teach sportsmanship skills to a child with autism. *Focus on Autism and Other Developmental Disabilities, 18,* 105–111. doi:10.1177/108835760301800204

Koegel, R. L., & Koegel, L. K. (1990). Extended reductions in stereotypic behavior of students with autism through a self-management treatment package. *Journal of Applied Behavior Analysis, 23,* 119–127. doi:10.1901/jaba.1990.23-119

Kozlowski, A., Wood, L., Gilligan, K., & Luiselli, J. K. (2009). Effects of nonverbal social disapproval on attention-maintained spitting and disruptive vocalizing in a child with autism. *Clinical Case Studies, 8,* 309–316. doi:10.1177/1534650109341840

Lang, R., O'Reilly, M., Sigafoos, J., Machalicek, W., Rispoli, M., Lancioni, G. E., … Fragale, C. (2010). The effects of an abolishing operation intervention component on play skills, challenging behavior, and stereotypy. *Behavior Modification, 34,* 267–289. doi:10.1177/0145445510370713

Lanovaz, M. J., & Sladeczek, I. E. (2011). Vocal stereotypy in children with autism: Structural characteristics, variability, and effects of auditory stimulation. *Research in Autism Spectrum Disorders, 5,* 1159–1168. doi:10.1016/j.rasd.2011.01.001

Lanovaz, M. J., Sladeczek, I. E., & Rapp, J. T. (2011). Effects of music on vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 44,* 647–651. doi:10.1901/jaba.2011.44-647

Liu-Gitz, L., & Banda, D. R. (2010). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. *Behavioral Interventions, 25,* 77–87. doi:10.1002/bin.297

Lorimer, P. A., Simpson, R. L., Smith Myles, B., & Ganz, J. B. (2002). The use of social stories as a preventative behavioral intervention in a home setting with a child with autism. *Journal of Positive Behavior Interventions, 4,* 53–60. doi:10.1177/109830070200400109

Love, S. R., Matson, J. L., & West, D. (1990). Mothers as effective therapists for autistic children's phobias. *Journal of Applied Behavior Analysis, 23,* 379–385. doi:10.1901/jaba.1990.23-379

Mace, F. C., Pratt, J. L., Prager, K. L., & Pritchard, D. (2011). An evaluation of three methods of saying “no” to avoid an escalating response class hierarchy. *Journal of Applied Behavior Analysis, 44,* 83–94. doi:10.1901/jaba.2011.44-83

Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: Reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism and Developmental Disorders, 30,* 599–606. doi:10.1023/A:1005695512163

Norris, C., & Dattilo, J. (1999). Evaluating effects of a social story intervention on a young girl with autism. *Focus on Autism and Other Developmental Disabilities, 14,* 180–186. doi:10.1177/108835769901400307

O’Reilly, M. F., Sigafoos, J., Lancioni, G., Rispoli, M., Lang, R., Chan, J., ... Langthorne, P. (2007). Manipulating the behavior-altering effect of the motivating operation. Examination of the influence on challenging behavior during leisure activities. *Research in Developmental Disabilities, 29,* 333–340. doi:10.1016/j.ridd.2007.06.004

Orr, T. J., Smith Myles, B., & Carlson, J. K. (1998). The impact of rhythmic entertainment on a person with autism. *Focus on Autism and Other Developmental Disabilities, 13,* 163–166. doi:10.1177/108835769801300304

Ozdemir, S. (2008). The effectiveness of social stories on decreasing disruptive behaviors of children with autism: Three case studies. *Journal of Autism and Developmental Disorders, 38,* 1689–1696. doi:10.1007/s10803-008-0551-0

Paterson, C. R., & Arco, L. (2007). Using video modeling for generalizing toy play in children with autism. *Behavior Modification, 31,* 660–681. doi:10.1177/0145445507301651

Peyton, R. T., Lindauer, S. E., & Richman, D. M. (2005). The effects of directive and nondirective prompts on noncompliant vocal behavior exhibited by a child with autism. *Journal of Applied Behavior Analysis, 38,* 251–255. doi:10.1901/jaba.2005.125-04

Rapp, J. T. (2007). Further evaluation of methods to identify matched stimulation. *Journal of Applied Behavior Analysis, 40,* 73–88. doi:10.1901/jaba.2007.142-05

Rapp, J. T., Vollmer, T. R., St. Peter, C., Dozier, C. L., & Cotnoir, N. M. (2004). Analysis of response allocation in individuals with multiple forms of stereotyped behavior. *Journal of Applied Behavior Analysis, 37,* 481–501. doi:10.1901/jaba.2004.37-481

Rehfeltdt, R. A., & Chambers, M. R. (2003). Functional analysis and treatment of verbal perseverations displayed by an adult with autism. *Journal of Applied Behavior Analysis, 36,* 259–261. doi:10.1901/jaba.2003.36-259

Rispoli, M., O'Reilly, M., Lang, R., Machalicek, W., Davis, T., Lancioni, G., & Sigafoos, J. (2011). Effects of motivating operations on problem and academic behavior in classrooms. *Journal of Applied Behavior Analysis, 44,* 187–192. doi:10.1901/jaba.2011.44-187

Roane, H. S., Falcomata, T. S., & Fisher, W. W. (2007). Applying the behavioral economics principle of unit price to DRO schedule thinning. *Journal of Applied Behavior Analysis, 40,* 529–534. doi:10.1901/jaba.2007.40-529

Roane, H. S., Piazza, C. C., Cercone, J. J., & Grados, M. (2002). Assessment and treatment of vocal tics associated with Tourette's syndrome. *Behavior Modification, 26,* 482–498. doi:10.1177/0145445502026004003

Ross, D. E. (2002). Replacing faulty conversational exchanges for children with autism by establishing a functionally equivalent response. *Education and Training in Mental Retardation and Developmental Disabilities, 37,* 343–362.

Rozenblat, E., Brown, J. L., Brown, A. K., Reeve, S. A., & Reeve, K. R. (2009). Effects of adjusting DRO schedules on the reduction of stereotypic vocalizations in children with autism. *Behavioral Interventions, 24,* 1–15. doi:10.1002/bin.270

Scattone, D., Wilczynski, S. M., Edwards, R. P., & Rabian B. (2002). Decreasing disruptive behaviors of children with autism using social stories. *Journal of Autism and Developmental Disorders, 32,* 535–543. doi:10.1023/A:1021250813367

Schumacher, B. I., & Rapp, J. T. (2011). Evaluation of the immediate and subsequent effects of response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 44,* 681–685. doi:10.1901/jaba.2011.44-681

Taylor, B. A., Hoch, H, & Weissman, M. (2005). The analysis and treatment of vocal stereotypy in a child with autism. *Behavioral Interventions, 20,* 239–253. doi:10.1002/bin.200

Thorp, D. M., Stahmer, A. C., & Schreibman, L. (1995). Effects of sociodramatic play training on children with autism. *Journal of Autism and Developmental Disorders, 25,* 265–282. doi:10.1007/BF02179288

Appendix C

Table C.1

Overview and Descriptive Statistics for Descriptive Variables and Potential Moderating Variables

Variable	Value	Descriptive statistics	Missing	*N*
Participant variables				
Age	Continuous	M = 10.1; SD = 7.80; range = 4–52	n = 1	N = 74
Gender	0 = male; 1 = female	n_0 = 53; n_1 = 20	n = 1	N = 74
VCB type	1 = stereotypical VCB; 2 = non- stereotypical VCB; 3 = combined VCB	n_1 = 45; n_2 = 22; n_3 = 8	n = 0	N = 75
Function of stereotypy if the VCB is stereotypical VCB[a]	1a = stereotypy automatically maintained; 1b = stereotypy reinforced by the environment; 1c = function not reported	n_{1a} =26; n_{1b} = 5; n_{1c} = 14	n = 0	N = 45
Additional language problem[b]	0 = no additional language problem; 1 = additional language problem	n_0 = 6; n_1 = 33	n = 35	N = 74
Additional diagnosis[c]	0 = normal hearing and vision according to school records; 1 = no any known sensory or physical deficits; 2 = Down syndrome; 3 = Tourette's syndrome; 4 = developmental delays; 5 = emotional disturbances and normal hearing and vision according to school records; 6 = major depression; 7 = pervasive developmental disorder; 8 = pervasive developmental disorder not otherwise specified and obsessive compulsive disorder; 9 = schizophrenia, developmental disabilities, non-organic psychosis; 10 = Waardenburg syndrome and severe hearing impairment; 11 = seizures; 12 = hypotonia, chronic otitis media, and congenital scoliosis	n_0 = 3; n_1 = 4; n_2 = 1; n_3 = 2; n_4 = 2; n_5 = 1; n_6 = 1; n_7 = 1; n_8 = 1; n_9 = 1; n_{10} = 1; n_{11} = 1; n_{12} = 1	n = 54	N = 74
Intellectual disability (level)[d]	0 = average intelligence; 1 = high-functioning; 2 = borderline intelligence; 3 = mental retardation; 4 = mild mental retardation; 5 = moderate mental retardation; 6 = severe mental retardation; 7 = profound mental retardation	n_0 = 4; n_1 = 3; n_2 = 2; n_3 = 2; n_4 = 3; n_5 = 4; n_6 = 4; n_7 = 1	n = 51	N = 74
Medication[e]	0 = no medication; 1 = medication	n_0 = 6; n_1 = 5	n = 63	N = 74
Intervention and context variables				
Pretreatment functional analysis	0 = no pretreatment functional analysis; 1 = pretreatment functional analysis	n_0 = 29; n_1 = 45	n = 0	N = 74

Table C.1 (Continued)

Variable	Value	Descriptive statistics	Missing	*N*
Type of pretreatment functional analysis[f]	0 = no functional analysis; 1 = indirect assessment with interview of scale; 2 = direct assessment, descriptive assessment, observation; 3 = experimental functional analysis; 4 = direct and indirect assessment; 5 = direct assessment and experimental functional analysis	$n_0 = 29$; $n_1 = 3$; $n_2 = 3$; $n_3 = 30$; $n_4 = 2$; $n_5 = 2$	$n = 5$	$N = 74$
Intervention type	1 = antecedent; 2 = consequence; 3 = combined intervention	$n_1 = 32$; $n_2 = 27$; $n_3 = 19$	$n = 0$	$N = 78$
Primary treatment setting	1 = community environment or treatment facility; 2 = home; 3 = school	$n_1 = 21$; $n_2 = 14$; $n_3 = 36$	$n = 3$	$N = 74$
Primary intervention agent	1 = investigator or experimenter; 2 = parent; 3 = teacher; 4 = therapist; 5 = staff or combinations of previous categories	$n_1 = 32$; $n_2 = 3$; $n_3 = 12$; $n_4 = 23$; $n_5 = 4$	$n = 0$	$N = 74$
Peers involved in treatment[g]	0 = no peers involved; 1 = peers involved	$n_0 = 72$; $n_1 = 2$	$n = 0$	$N = 74$
Design	1 = AB-design; 2 = reversal design; 3 = multiple baseline design; 4 = alternating treatments design	$n_1 = 7$; $n_2 = 25$; $n_3 = 20$; $n_4 = 22$	$n = 0$	$N = 74$
Duration of treatment (in weeks)[h]	Continuous	$M = 6.8$; $SD = 11.74$; range = 1–84	$n = 22$	$N = 74$
Frequency of treatment (sessions/week)[i]	Continuous	$M = 7.2$; $SD = 9.41$; range = 1.5–53.63	$n = 25$	$N = 74$
Treatment datapoints	Continuous	$M = 37.6$; $SD = 48.59$; range = 3–280	$n = 0$	$N = 74$
Study variables				
Publication year	Continuous	$M = 2005.4$; $SD = 5.44$; range = 1990–2011	$n = 0$	$N = 52$
Study quality	Continuous	$M = 7.6$; $SD = 1.06$; range = 6–10	$n = 0$	$N = 52$

Variables	Correlation
Duration of treatment and Frequency of treatment	$r = .62$
Duration of treatment and Treatment datapoints	$r = .32$
Frequency of treatment and Treatment datapoints	$r = .22$
Study quality and Publication year	$r = .05$

Notes. *M* = mean; *SD* = standard deviation; VCB = vocal challenging behaviour; *N* = total number of units; *n* = number of units; *r* = Pearson product-moment correlation coefficient.

[a, c, f, g]Function of stereotypy if the VCB is stereotypical VCB, additional diagnosis, type of pretreatment functional analysis, and involvement of peers were only used descriptively.

[b, d, e, h, i]The variables additional language problem (47.3% missing), intellectual disability (level) (68.9% missing), medication (85.1% missing), duration of treatment (29.7% missing), and frequency of treatment (33.8% missing) had a lot of missing data. Therefore, we decided to code the level of intellectual disability in a very detailed way. In addition, we presented these five variables in Table C.1 and excluded them from further analyses. Hence, they are not included in Figure D.1 and in Table 1 and Table 2.

Appendix D

Eleven testable potential moderators:
Age, gender, VCB type, FA, intervention type, primary treatment setting, primary intervention agent, design, treatment datapoints, publication year, and study quality.

↓

I: Four-variable model with the "standard" meta-analytical variables:

Age, $Z = 0.60$, $p = .5459$, gender, $Z = -0.71$, $p = .4806$, publication year, $Z = 0.23$, $p = .8193$, and study quality, $Z = 1.09$, $p = .2770$.

↓

II: Seven separate five-variable models with the four "standard" meta-analytical variables (ns) and …

… FA	… primary intervention agent	… design	… treatment datapoints
Data-analytical problems: "Too many likelihood evaluations"	$Z = 0.24$, $p = .8113$ *AIC* = 11142.7 *BIC* = 11142.7 *AICC* = 11156.1	$Z = -1.12$, $p = .2623$ *AIC* = 11140.0 *BIC* = 11140.0 *AICC* = 11153.4	$Z = 0.38$, $p = .7020$ *AIC* = 11156.8 *BIC* = 11156.8 *AICC* = 11170.2

… VCB type	**… intervention type**	**… primary treatment setting**
***EME* = 0.72, *Z* = 3.24, *p* = .0012**	***EME* = –0.72, *Z* = –3.50, *p* = .0005**	***Z* = 0.68, *p* = .4996**
***AIC* = 11090.8**	***AIC* = 11085.2**	***AIC* = 10930.0**
***BIC* = 11090.8**	***BIC* = 11085.2**	***BIC* = 10930.0** } **Best**
***AICC* = 11104.2**	***AICC* = 11098.5**	***AICC* = 10943.1** } **fit indices**

↓

III: Six-variable model with the "standard" meta-analytical variables (ns), VCB type, and intervention type

VCB type, $Z = 2.84$, $p = .0045$, and intervention type, $Z = -3.56$, $p = .0004$
AIC = 11040.2
BIC = 11040.2
AICC = 11053.6

↓

IV: Five separate seven-variable models with the "standard" meta-analytical variables (ns), VCB type **, intervention type *, and …**

… primary intervention agent	… design	… treatment datapoints
$Z = 0.33$, $p = .7387$ *AIC* = 11043.4 *BIC* = 11043.5 *AICC* = 11056.8**.**	$Z = -1.25$, $p = .2098$ *AIC* = 11041.6 *BIC* = 11041.7 *AICC* = 11055.0	$Z = 0.23$, $p = .8152$ *AIC* = 11057.9 *BIC* = 11058.0 *AICC* = 11071.3

… primary treatment setting	... FA
***Z* = 0.80, *p* = .4258**	*EME* = –1.07[a], $Z = -1.91$, $p = .0558$
***AIC* = 10832.0** } **Best fit indices**	*AIC* = 11036.7
***BIC* = 10832. 0** } **→ Definitive Model 2**	*BIC* = 11036.7
***AICC* = 10845.1** }	*AICC* = 11050.1

Further studying FA: One-variable model
FA, *EME* = 1.30[a], $Z = 2.94$, $p = .0033$

Further studying FA: Four two-variable models

FA and age	Data-analytical problems: "Too many likelihood evaluations"	FA, *EME* = 1.19[a], $Z = 2.71$, $p = 0.0067$, and gender
FA and publication year		FA, *EME* = 1.21[a], $Z = 2.64$, $p = .0083$, and study quality

Further studying FA: Two six-variable models with the "standard" meta-analytical variables (ns), …
… VCB type, and FA→ Data-analytical problems: "Too many likelihood evaluations"
… intervention type, *EME* = –0.82, $Z = -4.11$, $p < .0001$, and FA, *EME* = –1.29[a], $Z = -2.52$, $p = .0119$

Notes. EME = estimated moderator effect; ns = nonsignificant.
[a]The sign of the estimated moderator effect of FA differs between models.
*$p < .05$; **$p < .01$; ***$p < .001$.

Figure D.1. Development of Model 2.