Behaviour Research and Therapy 120 (2019) 103444

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Functional Behavioral Assessment-based interventions on adults' delusions, hallucinations and disorganized speech: A single case meta-analysis

María Xesús Froján-Parga, Miguel Núñez de Prado-Gordillo, Alejandra Álvarez-Iglesias, Jesús Alonso-Vega*

Universidad Autónoma de Madrid, Spain

ARTICLE INFO

Keywords:
Severe mental disorders
Psychological treatments
Functional behavioral assessment
FBA-based interventions
Single case desing
Metaanalysis

ABSTRACT

The Functional Behavioral Assessment (FBA) approach involves the use of single-case designs (SCD) to study the problem behavior-environment contingencies and conduct interventions that consider this functional relationship. Although this approach has been considered an evidence-based practice (EBP) for the treatment of several psychological problems, no meta-analytic studies of FBA-based interventions on delusions, hallucinations and disorganized speech -commonly operationalized as "atypical vocalizations"- have been carried out. Therefore, the purpose of this study was to review and synthesize the results of FBA-based interventions on adults' atypical vocalizations. We conducted a systematic review and a multi-level meta-analysis of these interventions, using a recently developed effect size estimator for SCD studies (i.e., log response ratio). All the studies that met our eligibility criteria provided evidence supporting the effectiveness of FBA-based interventions on atypical vocalizations, with an overall average effect size of a 72% reduction. Both the publication year and the methodological quality were found to be significant moderators. Despite some methodological limitations, we can conclude that FBA-based interventions are effective to reduce atypical vocalizations. The implications of these results could be of interest for the mental health community.

1. FBA-based interventions on adults' delusions, hallucinations and disorganized speech: a single case meta-analysis

Delusions, hallucinations and disorganized speech are symptoms of different mental illnesses and disorders as a schizophrenia, bipolar disorder or some types of dementia (APA, 2013). These phenomena are the source of serious problems in the functioning of personal, social and work daily life. In addition, they generate large economic and social costs (Whiteford et al., 2013). Although many efforts have been invested in dealing with these problems (Lutgens, Gariepy, & Malla, 2017; Skelton, Khokhar, & Thacker, 2015; Turner, van der Gaag, Karyotaki, & Cuijpers, 2014), providing an effective, efficient and ethically-informed treatment is still one of the main unresolved objectives of psychiatry and psychology (National Institute for Health and Care Excellence, 2014).

In this regard, the functional approach to psychology might offer some interesting insights. Within this approach, delusions, hallucinations and disorganized speech have traditionally been operationalized in terms of bizarre or atypical vocalizations (Sturmey, Ward-Horner, Marroquin, & Doran, 2007). A large body of evidence from the behavior analysis literature has shown that atypical vocalizations are behavioral responses influenced by environmental contingencies (Layng & Andronis, 1984; Mace, 1994; Mace, Lalli, & Lalli, 1991; Wong, 2014). These studies assess the contextual variables that maintain this kind of problems to produce significant clinical changes by directly modifying their maintaining variables (Mace, 1994).

Moreover, Functional Behavioral Assessment (FBA) is a pretreatment ideographic set of assessments which aim is to identify variables associated with the occurrence of a specific behavior, in order to develop an idiosyncratic intervention aimed at promoting behavioral changes (Iwata et al., 1994). There are three types of FBA: indirect (e.g. interviews), descriptive (e.g. direct observations) and experimental (e.g. manipulation of contextual variables). All of them have proven useful to determine the variables that maintain different problem behaviors (Beavers, Iwata, & Lerman, 2013). In addition, some assessment procedures can be modified to overcome the limitations intrinsic to a specific kind of problem, like psychotic behavior (Sturmey et al., 2007).

Therefore, FBA-based interventions are interventions guided by the results of a previous FBA. They typically employ a single-case design (SCD), which allows for the variables that control problem behavior to

* Corresponding author. Facultad de Psicología, C/ Ivan P. Pavlov, 6, 28049, Madrid, Spain.
E-mail address: jesus.alonsov@uam.es (J. Alonso-Vega).

https://doi.org/10.1016/j.brat.2019.103444
Received 22 November 2018; Received in revised form 29 June 2019; Accepted 26 July 2019
Available online 27 July 2019
0005-7967/

be detected and manipulated. These interventions are thus designed attending to the function of problem behavior (i.e., why does it occur) and not to its topography. They might also be more ecologically valid, since they can facilitate the generalization of clinical changes across different contexts. The more the FBA conditions resemble the natural circumstances in which the problem behavior occur, the more likely the successful achievement of a stable behavioral change will be (Hurl, Wightman, Haynes, & Virués-Ortega, 2016).

Since FBA-based interventions were first formally established in institutional contexts (Carr, 1977; Iwata, Dorsey, Slifer, Bauman, & Richman, 1994), the major developments and applications have occurred in educational settings and, specifically, in the intervention on children with developmental problems (Madden, Dube, Hackenberg, Hanley, & Lattal, 2013). However, the very first steps towards the development of formal FBA-based interventions were taken in institutionalized patients diagnosed with schizophrenia or other severe mental problems (Ayllon & Michael, 1959; Lindsley, 1956). These studies were the first to demonstrate how the behavior of people diagnosed with schizophrenia could be maintained and modified according to the principles of operant behavior (Ferster & Skinner, 1957).

The functional definition of hallucinations, delusions and disorganized speech has allowed for the establishment of a whole research field focused on the development of interventions aimed at achieving a significant clinical change in atypical verbal behaviors traditionally associated with schizophrenia (Rosenfarb, 2013). The main advances in this research field could be summarized in four key points:

1) Pathognomonic symptoms of schizophrenia (e.g., hallucinations, delusions, etc.) or other serious mental illnesses are problems that can be successfully treated by managing the contingencies that maintain these behaviors (Burns, Heiby, & Tharp, 1983)
2) Functional analysis can be used to establish these maintaining contingencies (Mace et al., 1991)
3) These contingencies can be managed and controlled through verbal interaction (Baruch, Kanter, Busch, & Juskiewicz, 2009)
4) Through this type of intervention, pharmaceutical spending is reduced (Markwick, Smith, & Mick, 2014).

Briefly, the key achievement of behavioral analysis in clinical contexts is that it has proven effective in changing user behaviors in a cost-efficient, non-invasive and idiosyncratic way (Madden et al., 2013).

According to APA Presidential Task Force (APA, 2006), systematic reviews and meta-analyses (e.g., Lutgens et al., 2017; Turner et al., 2014) have been established as a useful tool to evaluate and compare the effects of different interventions and thus determine if a given procedure can be regarded as an Evidence-Based Practice (EBP). However, few publications have focused on studying the available evidence of FBA-based interventions. Perhaps this lack of studies is due to common misunderstandings about the SCD used in FBA scientific literature (see Shadish, 2014). However, due to the efforts made to improve the mathematical models that allow for a more adequate synthesis of SCD studies and to the renewed interest in person-centered practice, these types of studies are increasingly being recognized as a valuable source of evidence for improving decision-making in health systems (Shadish, Hedges, & Pustejovsky, 2014). For example, Common, Lane, Pustejovsky, Johnson, and Johl (2017) carried out a meta-analysis of FBA–based interventions for students with or at-risk of high-incidence disabilities, showing that FBA–based interventions could be determined as an EBP following the Standards for EBP (APA, 2006; Shadish, 2014). In addition, Hurl et al. (2016) performed a meta-analysis of studies comparing FBA-based interventions with non-FBA-based interventions. They found that, while the former had a large effect on the reduction of problem behavior, the latter had no effect when compared to no intervention. They also showed that the effect of FBA-based interventions on appropriate behaviors was four times greater than the effect found in non-FBA-based interventions.

However, research on FBA-based interventions on problem behaviors other than developmental or school-related problems is still scarce. Regarding FBA-based interventions on delusions, hallucinations and disorganized speech, different narrative reviews have been published (Layng & Andronis, 1984; Mace, 1994; Mace et al., 1991; Travis & Sturmey, 2008; Wong, 2014) but no systematic reviews nor meta-analytic syntheses have been carried out. Therefore, the main objective of this paper is to review and synthetize the published evidence of FBA-based interventions on atypical vocalizations.

2. Method

2.1. Literature search

The research question in plain terms was: Are FBA-based interventions effective in treating adults' delusions, hallucinations and disorganized speech? This question was elaborated considering the strategy PICOS commonly used to identify components of clinical evidence for systematic reviews in evidence-based practice and is endorsed by the Cochrane Collaboration (Higgins et al., 2011).

Consequently, we reviewed different SCD studies that reported outcomes of FBA-based interventions on adults' atypical vocalizations. The literature search was conducted using the following databases: PsycInfo, PubMed, Web of Science, and Open Gray. Common search terms employed were *functional analysis, hallucinatory speech*, *delusional statements* and *disorganized speech*. Only papers published in English or Spanish were included, both because the vast majority of scientific literature is published in these languages and because these were the only languages well known by the authors. No restrictions on the publication date were applied. Preliminary searches started in January 2018 and formal screening of search results against eligibility criteria was carried out in February 2018. This work was carried out following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses standards (Moher, Liberati, Tetzlaff, & Altman, 2009) and reporting standards of the Meta-Analysis Reporting Standards (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). The search strategy can be found in the following link: https://osf.io/7vzda/?view_only=92e057b116bc4297a70557c7420f680d. In addition, the search protocol was pre-registered in PROSPERO.

Studies were considered for inclusion if they met the following criteria: single-case design and FBA-based interventions for hallucinatory speech (e.g., verbal responses to not present stimuli), delusional speech (e.g., obviously false statements) or disorganized speech (e.g., stereotyped or repetitive verbal responses) in adult participants. All cases were included regardless of the mental disorder diagnosis or concurrent pharmacological treatment. Studies that did not conduct a functional assessment or did not perform an intervention on an adult were excluded. A total of 213 studies were retrieved from the database searches and, finally, 23 SCD studies met inclusion criteria and were included in the review. Fig. 1 shows the study selection process flux diagram.

2.2. Variables and data extraction procedure

The data extraction of the selected studies was focused on the following variables: patient information (age, gender, mental disorder diagnosis), problem behavior information (behavioral topography and behavior function), intervention characteristics (functional assessment method, intervention technique, duration, intervention setting, concurrent pharmacological treatment and pharmacological treatment changes), methodological characteristics (single-subject experimental design) and intervention outcomes (behavioral direct measures in baseline, post-treatment data and follow-up data of problem behaviors and appropriate behaviors). To extract the results of each intervention we used the WebPlotDigitizer software (Rohatgi, 2018), as

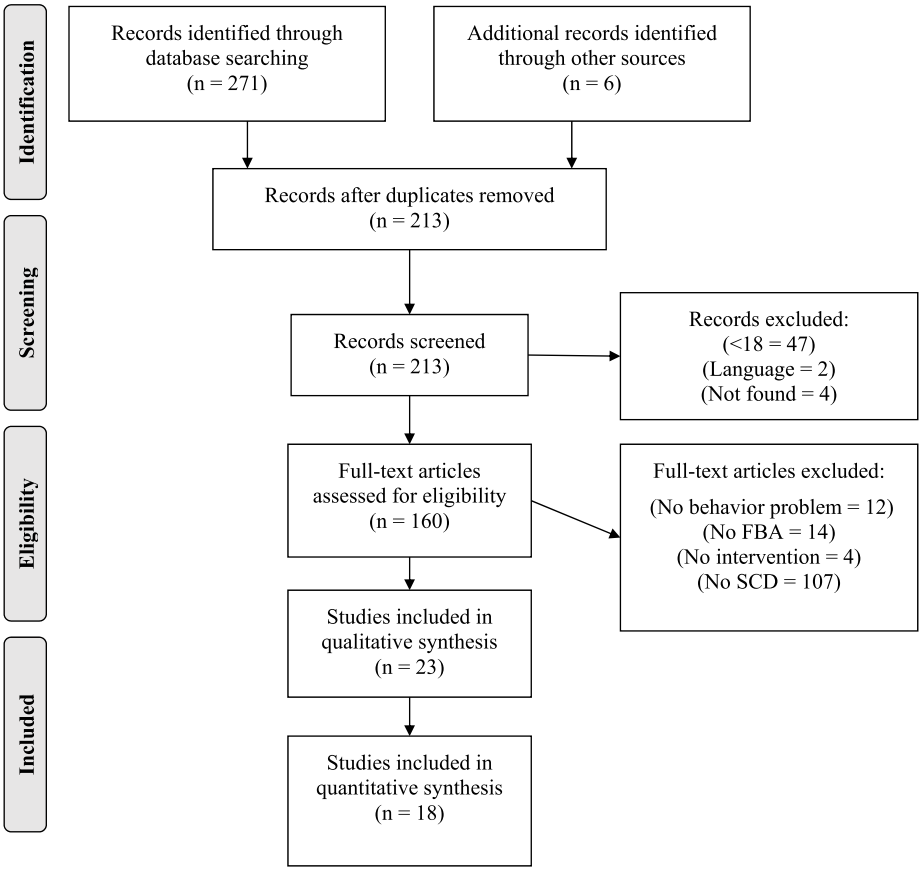

Fig. 1. Flux diagram.

recommended by Moeyaert, Maggin, and Verkuilen (2016). This software allowed us to extract each individual data point of the behavioral measures reflected in the graphs provided by the included works.

The literature search process inter-rater reliability was separately calculated as follows. Firstly, screening inter-rater reliability was calculated as the percentage of times that both raters independently evaluated a publication as either eligible or non-eligible for inclusion; it was 94%. Secondly the inter-rater reliability of intervention and methodological study characteristics was calculated as the percentage of data variables within each study that were exactly rated by both researchers. All studies were evaluated by one researcher, while the other reviewed 1 of every 3 studies selected at random; the inter-rater reliability at this phase was 100%. Finally, data extraction inter-rater reliability –calculated as the percentage of data extracted from one intervention per case within each study that were equally rated by both researchers, with a 1% margin of error-, was conducted in the same way. Again, inter-rater reliability at this phase was 100%. All disagreements were settled by consensus with a third researcher.

2.3. Quality analysis

We conducted a quality analysis to determine the methodological quality of the study sample. Two researchers independently conducted a quality analysis of each intervention within each SCD study, following the criteria of the What Works Clearinghouse (WWC) SCD Panel (Kratochwill et al., 2010). We chose to follow the WWC SCD recommendations because it specifies in great practical detail how to conduct a comprehensive analysis of the key methodological characteristics of a good single case design. In addition, it allowed us to classify each intervention within each reviewed study depending on a) whether it met the WWC standards, it met them with reservations or it did not meet them; and b) whether its visual analysis showed strong, moderate or no evidence of a causal relation between the intervention and the observed changes in the target behavioral outcome. To quantify the results of the quality analysis, each intervention was assigned a score from 0 to 6 following the quality index guidelines described in Hurl et al. (2016), p. 6 = interventions that meet the WWC standards and present strong evidence of a causal relation; 5 = meet the standards and present moderate evidence; 4 = meet the standards with reservations and present strong evidence; 3 = meet the standards with reservations and present moderate evidence; 2 = meet the standards and present no evidence; 1 = meet the standards with reservations and present no evidence; and 0 = do not meet the standards.

The quality analysis inter-rater reliability was calculated as the percentage of included interventions which methodological characteristics and visually-inspected evidence for a causal relation were rated the same by both researchers. As with the data extraction process, one researcher rated all studies and the other was randomly assigned the 30%. The quality analysis inter-rater reliability was 100% for methodological characteristics and 90% for the visual analysis of the intervention outcome. All disagreements were settled by consensus.

2.4. Effect size analysis

Among the different effect size estimators found in the literature, we chose the Log Response Ratio (LRR; LRRd for a decrease size effect and LRRi for an increase size effect), a recently developed effect size estimator for single-case studies of free-operant behavior (Pustejovsky, 2015, 2018). It was chosen for its statistical properties and for its adequacy to the characteristics of our study sample (see Pustejovsky, 2015, 2018; Zimmerman et al., 2018). In addition, in contrast to other parametric effect sizes, this within-case parametric effect size estimator can be obtained for each case within a SCD study, thus allowing for a quantitative synthesis of single-case interventions to be performed

(Pustejovsky, 2015, 2018). The LRR effect size parameter is defined as:

$$\psi = \ln\left(\frac{\mu_B}{\mu_A}\right),$$

where ln() stands for the natural logarithm function, μ_A stands for the baseline mean level and μ_B stands for the treatment mean level. However, this effect size parameter implies a series of assumptions that may not be realistic given the characteristics of many SCD studies. Thus, a series of bias corrections are needed to achieve a proper effect size estimator. Specifically, the problem of auto-correlation (i.e., non-independent sampling of the outcome measure) could yield biased LRR sampling variances (Pustejovsky, 2018a, 2018b, 2018c). A method to control for this problem is outlined below.

Although the main target of our review were the interventions on atypical vocalizations (i.e., LRRd), the LRR was also calculated for interventions on appropriate behavior (i.e., LRRi), when this measure was included. Since 5 studies (Alumbaugh, 1971; Ayllon & Michael, 1959; Nydegger, 1972; Slade, 1972; and; Vandbakk, Arntzen, Gisnaas, Antonsen, & Gundhus, 2012) did not assessed behavioral outcomes through direct observation or did not report the results across each intervention phase, the LRRd was estimated from the data of 19 cases (18 studies), out of the 23 studies that first met the inclusion criteria. Out of those 19 cases, only 11 included measures of appropriate behavior. Since many cases included several interventions, we only calculated the LRR for the intervention that showed the highest quality analysis index. All calculations were done with the software R v3.0.5 (R Core Team, 2018), specifically the SingleCaseES package (Pustejovsky, 2018a, 2018b, 2018c).

2.5. Quantitative synthesis

We then proceeded to conduct a quantitative synthesis of the results of the 19 FBA-based interventions for atypical vocalizations. The WWC SCD Panel stablishes that a 5-3-20 threshold must be met before the results of a given set of SCD studies can be summarized, so that it must include: a) at least 5 studies that meet the WWC standards with or without reservations; b) studies carried out by at least 3 different research teams from three different institutions with no overlapping authorship; c) a total number of cases of at least 20. Although our study sample met a) and b), it did not meet c), given that we only had 19 case interventions. Still, since the characteristics of our study sample were quite close to meet the WWC 5-3-20 threshold, we deemed it appropriate to conduct an exploratory quantitative synthesis.

Prior to the analysis, we screened for standardized z values larger than 3.29 or smaller than −3.29 to control for the presence of potential outliers that could influence the overall average outcome (Assink & Wibbelink, 2016). We then conducted a random effects multilevel meta-analysis (Assink & Wibbelink, 2016; Konstantopoulos, 2011) using the metafor package in R (Viechtbauer, 2010). The random effects approach to meta-analysis was deemed appropriate given the considerable degree of heterogeneity of our study sample. On the other hand, the multilevel meta-analytic model was deemed appropriate since it adds random effects at both the case level and the study level, thus accounting for the problem of correlation between effects within the same study (Konstantopoulos, 2011; Pustejovsky, 2018a, 2018b, 2018c).

Therefore, the main possible sources of heterogeneity in our meta-analytic model were three: sampling variance (i.e., heterogeneity due to the recording of the target behavior), within-study variance (i.e., heterogeneity due to the differences among cases within a given study) and between-study variance (i.e., heterogeneity due to the differences among studies). Estimates of within-study variance (ω^2) and between-study variance (τ^2) were obtained through a restricted maximum likelihood method (Pustejovsky, 2018a, 2018b, 2018c; Viechtbauer, 2010). However, the estimation of both variance components is dependent on the accuracy of the sampling variance, which in the case of SCD studies might be affected by the abovementioned problem of auto-correlation (Hedges, Tipton, & Johnson, 2010). A procedure of robust variance estimation with small-sample corrections was applied through the clubSandwich package in R (Pustejovsky, 2018a, 2018b, 2018c), to obtain robust ω^2 and τ^2 estimates even in the presence of auto-correlation (Hedges et al., 2010; Tipton, 2015). In addition, we performed two separated log-likelihood-ratio tests to determine whether the within-study and between-study variance components were significant (Assink & Wibbelink, 2016). However, given the potential threat that the log-likelihood-ratio tests would not be significant due to the small number of effect sizes, we also calculated the percentage of the total amount of heterogeneity located at each level (sampling variance, case-level and study-level). We applied the 75% rule described in Assink and Wibbelink (2016), which states that if the sampling variance does not account for at least the 75% of the total amount of variance, then heterogeneity can be regarded as substantial and moderator analyses should be performed.

Finally, to account for the amount of heterogeneity due to the differences in the abovementioned study characteristics, we performed a mixed effects multilevel meta-analysis. We first analyzed every potential moderator separately to detect which were significant. Once single significant moderators were detected, a mixed effects multilevel meta-analysis with multiple moderators was performed.

3. Results

3.1. Characteristics of the study sample

A total of 23 studies (see Table 2) were included in this review. Tables 1 and 2 present the study characteristics and participant information. These studies report the results of FBA-based interventions for atypical vocalizations of 24 adult participants (54.2% female). One of the included studies reported two cases. All participants were adults (mean = 35.5) no older than 65 years. Most participants had multiple diagnosis (41.7%). The most frequent single diagnosis was schizophrenia (29.2%). 41.7% of the participants presented multiple problem behavior topography; delusional plus hallucinatory speech was the most common one (33.3%). In addition, attention was the most common function of problem behavior (58.33%).

The most used FBA method was the experimental one (41.1%). In 5 studies, two FBA-methods were used. The AB^K design was the most used in the different interventions (91%); only two studies employed a multiple baseline design. Interventions generally included several behavior modification techniques (62.5%). The most common combined intervention was the differential reinforcement of appropriate behaviors plus extinction of the problem behaviors (50%). In addition, nearly half the interventions used natural change agents (45.8%).

Regarding the methodological quality, the different interventions included in each case (n = 48) showed great differences in their associated quality analysis indexes. We rated 29 interventions (60.4%) intended to decrease atypical vocalizations (extracted from the total 24 cases) and 19 interventions (40.6%) intended to increase appropriate vocalizations (extracted from 17 cases). 12 interventions on atypical vocalizations (41.7%), extracted from 10 different cases, demonstrated acceptable methodological quality rates (i.e., QA index ≥ 3). On the contrary, only 5 FBA-interventions on appropriate behavior showed acceptable methodological quality indexes. Many of the 0 scores were due to the employment of a pre-experimental design (e.g., did not have a minimal AB^2 or ABAB design).

3.2. Intervention outcomes and meta-analysis

Table 3 reports the outcomes of the interventions included in the meta-analysis in terms of percentage decrease of atypical vocalizations and percentage increase of appropriate behavior, whenever this last

Table 1
Participants, setting and study characteristics.

	Included Cases
	n (%)
Gender	
Female	13 (54.2)
Male	11 (45.8)
Age	
Early adults (18–25 years)	6 (25)
Adults (26–65 years	17 (70.8)
Unspecified	1 (4.2)
Diagnosis	
Autism spectrum disorder	1 (4.2)
Moderate intellectual disability	2 (8.3)
Severe intellectual disability	1 (4.2)
Schizophrenia	7 (29.2)
Traumatic brain damage	1 (4.2)
Multiple	10 (41.7)
Unspecified	2 (8.3)
Behavior Topography	
Delusional speech	4 (16.7)
Hallucinatory speech	5 (20.8)
Disorganized speech	5 (20.8)
Multiple	10 (41.7)
Delusional + Hallucinatory	8 (33.3)
Delusional + Disorganized	2 (8.3)
FBA method	
Descriptive	3 (12.5)
Experimental	10 (41.1)
Indirec	6 (25)
Multiple	5 (20.8)
Indirect + Experimental	3 (12.5)
Indirect + Descriptive	1 (4.2)
Descriptive + Experimental	1 (4.2)
Behavior Function	
Positive reinforcement	15 (62.5)
Attention	15 (62.5)
Negative reinforcement	1 (4.2)
Escape from aversive stimuli	1 (4.2)
Respondent conditioning	2 (8.3)
Multiple	5 (20.8)
Escape/Attention	3 (12.5)
Escape/RC	1 (4.2)
Automatic reinforcement[1]/Tangibles	1 (4.2)
Not found	1 (4.2)
Intervention Technique	
DR	16 (66.7)
+ Ex	12 (50)
+ Inst	1 (4.2)
+ SD	1 (4.2)
NCR	4 (16.7)
TO	2 (8.3)
+ Inst	1 (4.2)
SD	2 (8.3)
Single Case Design	
$AB^{\geq 2}$	14 (58.3)
$AB^{< 2}$	8 (33.3)
Multiple Baseline	2 (8.3)

Notes. Intervention technique: DR = differential reinforcement; Ex = Extinction; Inst = Instructions; NCR = non-contingent reinforcement; TO = time out; and SD = systematic desensization. Single Case Design: $AB^{\geq 2}$ = SCD with at least two baseline and two treatment phases; $AB^{< 2}$ = SCD with less than two baseline or two treatment phases. [1] Unspecified if positive or negative.

measure was included. These percentage measures were obtained by a transformation of the LRRd and LRRi[1] estimates. Additional information regarding the generalization or maintenance in time (i.e., follow-up) of reported results on atypical vocalizations is also displayed here. Only 36.8% of interventions reported some kind of generalization test (many of them were simply anecdotal), out of which 71.4% reported positive generalization results. Even fewer interventions reported some kind of follow-up information (21.1%), out of which 75% reported positive maintenance results.

Table 4 shows a summary of descriptive statistics, LRRd and its associated standard error for each intervention phase. Since none of the standardized z values associated to each LRRd estimate were larger than 3.29 or smaller than −3.29, all the reviewed interventions were included in the meta-analysis.

Fig. 2 presents the forest diagram of multivariate meta-analysis of random effects of selected interventions in atypical vocalizations. Interventions are ordered from greater to smaller effect size and the confidence interval of each of the interventions is represented. In this figure we can see how all the interventions have shown a significant effect size in which the value 0 - which means no effect - is not within their confidence interval.

Table 5 includes a summary of the results of both the random effects multilevel meta-analysis (Model I) and the mixed effects multilevel meta-analysis (Model II). Across the 19 cases, the overall average effect size was −1.26 ($p < .001$), 95% CI: [-0.959, −1.56], which approximately corresponds to a percentage decrease in atypical vocalizations of 72%, 95% CI: [62%, 79%]. The associated robust standard error was 0.142. The robust within-study variance component was $\omega^2 = 0.001$. This result was expected given the low number of studies with more than one case. On the other hand, the robust between-study variance estimator was $\tau^2 = 0.309$. This result could indicate a high degree of unaccounted heterogeneity in effects across studies. However, none of the two separated log-likelihood-ratio tests for each variance component yielded significant results ($p > .05$). Nonetheless, since the percentage amount of the total variance that could be attributed to the sampling variance level was just an 8%, we decided to follow the 75% rule described in Assink and Wibbelink (2016) and keep the multi-level meta-analytic model while checking for possible moderators.

Then, we conducted a mixed effects multilevel meta-analysis to detect potential moderators that could partially account for the unexplained heterogeneity. The potential continuous moderators considered were the publication year and the participant's age. The potential categorical moderators considered were: participant's gender (male or female), type of functional assessment (indirect, direct, experimental or mixed), intervention technique (differential reinforcement, non-contingent reinforcement or time-out), behavior topography (hallucinations, delusions, disorganized speech or mixed), diagnosis (schizophrenia, multiple or other), diagnosis nature (developmental vs. non-developmental condition), recording procedure (event counting, continuous recording or partial interval recording) and the quality analysis index (QA-0, QA-3 or QA-5).

The omnibus tests for each potential moderator yielded only two significant moderators: the publication year ($F(1, 17) = 4.497$, $p = .034$) and the quality analysis index ($F(2, 16) = 9.393$, $p = .009$). The publication year regression coefficient (−0.020, p = .036) showed that the more recent the interventions were, the larger (i.e., more negative) the effect size was. As to the quality analysis index, the average effect size was significant for all levels (QA-0 = −1.05, $p < .001$; QA-3 = −1.07, $p < .01$; QA-5 = 1.95, $p < .01$), although the QA-5 group showed a significantly larger effect size than both the QA-0 group (estimated difference = −0.90, $p < .05$) and the QA-3 group (estimated difference = −0.88, $p < .05$). No significant differences were

[1] As it was explained above, we could not carry out a quantitative synthesis of the results of the interventions aimed at increasing alternative behavior due to an insufficient number of case interventions including this kind of measure (only 11). Therefore, the LRRi for those interventions is not reported in Table 4, because it was not finally employed for conducting a quantitative synthesis.

(*footnote continued*)
However, we employed it to obtain a more precise measure of the percentage increase of appropriate behavior obtained in each of those interventions, which appears in Table 3.

Table 2
Intervention characteristics.

Study	Problem Behavior	Function	Design	Quality Index[1]	Age	Diagnosis	FBA	FBA-based intervention[2]	Use of natural change agents
Alumbaugh (1971)	H	RC	$AB^{<2}$	0	53	None	IND	SD	Yes
Anderson and Alpert (1974)	H	Attention	$AB^{\geq 2}$	0	26	SCHZ	DES	DRO	No
Arntzen, Tonnessen, and Brouwer (2006)	D/DS	Attention	$AB^{<2}$	0	44	Multiple	IND/EX	DRA/Extinction	No
Ayllon and Haughton (1964)	D	Attention	$AB^{<2}$	0	47	SCHZ	EX	DRA/Extinction	Yes
Ayllon and Michael (1959)	D	Attention	$AB^{<2}$	0	*	SCHZ	IND	DRA/Extinction	Yes
Carr and Britton (1999)	DS	Attention	MB	0	32	MID	IND/EX	NCR	No
Davis, Wallace, Liberman, and Finch (1976)	D/H	Attention	$AB^{\geq 2}$	0	33	SCHZ	IND	TO/Instructions	Yes
DeLeon, Arnold, Rodriguez-Catter, and Uy (2003)	DS	Attention	$AB^{\geq 2}$	5	21	Multiple	EX	DRA/Extinction	No
Dixon, Benedict, and Larson (2001)	D/H	Attention	$AB^{\geq 2}$	3	25	Multiple	EX	DRA/Extinction	No
Haynes and Geddy (1973)	H	*	$AB^{\geq 2}$	0	45	SCHZ	DES	TO	Yes
Horner, Albin, and Mank (1989)	DS	Attention	$AB^{\geq 2}$	3	26	SID	IND/EX	DRA/Extinction	No
Jimenez, Todman, Pérez, Godoy, and Landon-Jimenez (1996)	H	Escape	$AB^{\geq 2}$	0	49	SCHZ	DES	DRO/Instructions	No
Lancaster et al. (2004)									
Participant 1	D/H	Attention	$AB^{\geq 2}$	3	56	Multiple	EX	NCR	No
Participant 3	D/H	Attention	$AB^{\geq 2}$	3	53	Multiple	EX	NCR	No
Mace and Lalli (1991)	D/H	Escape/ Attention	$AB^{\geq 2}$	5	46	MID	EX/DES	NCR	Yes
Mace, Webb, Sharkey, Mattson, and Rosen (1988)	D/H	Escape/ Attention	$AB^{\geq 2}$	5	29	Multiple	EX	DRA/Extinction	Yes
McDonough, Johnson, and Waters (2017)	D	Escape/ Attention	$AB^{<2}$	0	45	Multiple	IND/DES	DRA/Extinction	Yes
Nydegger (1972)	D/H	Escape/RC	$AB^{<2}$	5	20	SCHZ	IND	DRA/SD	Yes
Rehfeldt and Chambers (2003)	DS	Attention	$AB^{\geq 2}$	3	23	Multiple	EX	DRA/Extinction	No
Slade (1972)	H	RC	$AB^{<2}$	5	18	None	IND	SD	No
Travis and Sturmey (2010)	DS	Attention	$AB^{\geq 2}$	3	26	TBD	EX	DRA/Extinction	Yes
Vandbakk et al. (2012)	D/DS	Auto/Tangible	$AB^{<2}$	0	24	ASD	IND	DRA	No
Wilder, Masuda, O'Connor, and Baham (2001)	D	Attention	$AB^{\geq 2}$	5	43	Multiple	EX	DRA/Extinction	No
Wilder, White, and Yu (2003)	D/H	Attention	MB	0	36	Multiple	EX	DRA/Extinction	Yes

Notes. Behavioral problem: H = hallucination; D = delusion; DS = disorganized speech. Function: RC = respondent conditioning; Auto = automatic reinforcement. Mental diagnostic: SCHZ = schizophrenia; MID = mild intellectual disorder; SID = severe intellectual disorder; TB = traumatic brain disorder; ASD = autism spectrum disorder. Behavioral technique: SD = systematic desensitization; DRO = differential reinforcement of other behaviors; DRA = Differential reinforcement of alternative behavior; TO = time out and NCR = non contingent reinforcement. Design: $AB^{\geq 2}$ = SCD with at least two baseline and two treatment phases; $AB^{<2}$ = SCD with less than two baseline or two treatment phases. [1] Single-case design quality index for the interventions on atypical vocalizations. [2] Only the intervention techniques included in the meta-analysis were included, except in the case of Alumbaugh (1971), Ayllon and Michael (1959), Nydegger (1972), Slade (1972) and Vandbakk et al. (2012). The interventions reported in these five studies were not included in the meta-analysis because they did not assess behavioral outcomes through direct observation or did not report the results across each intervention phase.

found between the QA-3 and the QA-0.

Subsequently, we performed a mixed effects multi-level meta-analysis with both moderators. The publication year was found to be only marginally significant (-0.021, $p < .1$), thus suggesting that the previously observed moderating effect of this variable was at least partially confounded with the moderating effect of the quality analysis index. The average effect size was significant for all levels of the quality analysis index (QA-0 = -1.17, $p < .01$; QA-3 = -0.93, $p < .01$; QA-5 = -1.94, $p < .01$), which corresponded to percentage decreases in atypical vocalizations of 69%, 95% CI: [46%, 82%] for the QA-0 group; 61%, 95% CI: [42%, 73%] for the QA-3 group; and 86%, 95% CI: [76%, 92%] for the QA-5 group. Again, the QA- 5 showed a significantly larger effect size than both the QA-0 (estimated difference = -0.78, $p < .01$) and the Q-3 groups (estimated difference = -1.01, $p < .01$), and no significant difference was found between these last two.

Finally, while the robust within-study variance estimator remained almost equal ($\omega^2 = 0.002$), the robust between-study variance estimator changed to $\tau^2 = 0.138$, thus indicating that both moderators accounted for the 55% of the between-study heterogeneity. However, the test for residual heterogeneity was still significant ($p < .001$), thus pointing at the possible influence of other variables not considered in our model on the FBA-intervention effectiveness.

4. Discussion

After analyzing the results, it can be concluded that all the included cases of FBA-based interventions for atypical vocalizations have proven to be effective at detecting the behavioral functions of this kind of problems and, consequently, implementing intervention techniques aimed at reducing them, with high overall percentage reduction results. However, included studies seldomly reported information regarding the generalization or maintenance in time of these outcomes (although, when they did, the results tended to be positive). In addition, in many studies the frequency of appropriate behavior also increased. Still, this last effect was not always observed, and it was highly variable. This high variability might be partially due to the high sensitivity of the LRRi estimate when the target behavior level approaches zero at any given intervention phase (Pustejovsky, 2018a, 2018b, 2018c). However, it might also be probably due to the fact that the main focus of the included FBA-based interventions was on the reduction of atypical vocalizations rather than the promotion of alternative behaviors. Thus, although we can be confident about the behavior reduction effects of this type of interventions, we cannot draw firm conclusions on their appropriate behavior increasing effects. Both issues (i.e., the relative lack of evidence on appropriate behavior increase and the lack of generalization and follow-up measures) should be taken into account by future clinical research on this area.

Regarding the results of the meta-analysis, the FBA-based

Table 3
Intervention outcomes.

Study[a]	Problem behavior reduction (%)	Appropriate behavior increases (%)	Reduction generalization/ Achievement[b]	Reduction maintenance (follow-up)/ Achievement[c]
Anderson and Alpert (1974)	50.1	*	Yes/Yes	No/*
Arntzen et al. (2006)	47.8	543.8	Yes/Yes	No/*
Ayllon and Haughton (1964)	60.4	237.1	No/*	No/*
Carr and Britton (1999)	59.7	*	No/*	No/*
Davis et al. (1976)	62.9	*	Yes/No	Yes/No (unspecified)
DeLeon et al. (2003)	89.5	*	No/*	No/*
Dixon et al. (2001)	56.1	268.5	No/*	No/*
Haynes and Geddy (1973)	37.5	*	No/*	No/*
Horner et al. (1989)	47.4	12.1	Yes/No	No/*
Jimenez et al. (1996)	87.4	*	No/*	Yes/Yes (6 weeks)
Lancaster et al. (2004)				
Participant 1	69.2	206.9	No/*	No/*
Participant 2	61.9	−42.8	No/*	No/*
Mace and Lalli (1991)	85.4	*	Yes/Yes	No/*
Mace et al. (1988)	78.1	112.3	Yes/Yes	No/*
McDonough et al. (2017)	91.1	*	No/*	Yes/Yes (2, 4 and 6 months).
Rehfeldt and Chambers (2003)	63.9	144	No/*	No/*
Travis and Sturmey (2010)	85.1	259.6	No/*	Yes/Yes (6 months, 1, 2 and 4 years)
Wilder et al. (2001)	91.5	96.1	No/*	No/*
Wilder et al. (2003)	73.6	32.5	Yes/Yes	No/*

[a] Only the outcomes of the 19 interventions included in the meta-analysis are reported.
[b] The achievement of the generalization of reduced problem behavior is reported dichotomously since not all interventions included specific generalization measures (e.g., some of them merely reported generalization in a narrative way or included mixed data from diverse treatment or extra-treatment settings). In addition, some studies reported specific generalization measures, but these were either: a) already included in the data employed for the calculation of the overall percentage decrease of problem behavior (when these measures were part of the intervention included in the meta-analysis); or b) were part of a separate intervention that was not included in the meta-analysis and, thus, was not quantitatively analyzed by the authors of the present study.
[c] The achievement of maintenance of reduced problem behavior at follow-up is reported dichotomously for not all interventions included specific measures (e.g., some of them merely reported maintenance in a narrative way). In addition, some studies reported specific follow-up measures, but these were screened out of the data gathering for the evaluation of intervention efficacy and were not quantitatively analyzed by the authors of the present study.

Table 4
Summary statistics and LLR effect size estimates for atypical vocalizations of included studies.

Study	Baseline phase			Treatment phase			Effect side	
	$\tilde{y}_A$	s_A	n	$\tilde{y}_A$	s_A	n	LRR	SE^R
Anderson and Alpert (1974)	52.310	5.786	8	25.8716	15.340	18	−0.695	0.145
Arntzen et al. (2006)	20.673	5.342	7	10.814	5.897	41	−0.649	0.129
Ayllon and Haughton (1964)	77.163	27.510	20	30.196	21.548	18	−0.927	0.186
Carr and Britton (1999)	98.815	2.450	19	39.634	29.491	61	−0.909	0.095
Davis et al. (1976)	50.395	25.286	7	18.843	22.854	79	−0.992	0.233
DeLeon et al. (2003)	16.677	2.033	22	0.176	0.172	16	−2.249	0.356
Dixon et al. (2001)	12.496	3.859	6	5.510	1.825	19	−0.823	0.147
Haynes and Geddy (1973)	47.258	5.843	13	29.393	13.657	22	−0.470	0.104
Horner et al. (1989)	1.6482	0.206	10	0.864	0.121	11	−0.641	0.149
Jimenez et al. (1996)	14.772	9.547	8	1.864	2.019	23	−2.07	0.321
Lancaster et al. (2004)								
Participant 1	49.262	6.985	9	15.069	6.754	12	−1.177	0.137
Participant 2	80.951	15.416	8	30.876	15.594	10	−0.953	0.173
Mace and Lalli (1991)	22.893	12.196	38	3.271	4.995	50	−1.926	0.232
Mace et al. (1988)	4.145	0.913	15	0.904	0.601	53	−1.520	0.107
McDonough et al. (2017)	51.514	19.084	11	4.113	9.331	20	−2.420	0.484
Rehfeldt and Chambers (2003)	27.518	10.062	11	9.871	7.055	20	−1.018	0.194
Travis and Sturmey (2010)	1.445	0.175	8	0.209	0.157	9	−1.902	0.253
Wilder et al. (2001)	31.773	25.106	11	2.555	4.622	20	−2.466	0.469
Wilder et al. (2003)	15.545	11.653	9	4.083	3.823	22	−1.331	0.263

interventions on atypical vocalizations show a high degree of behavioral change, with associated percentage decreases of problem behavior ranging from 62% to 79%. This suggests that the functional analysis of behavior is a reliable assessment tool to guide the treatment of atypical vocalizations and consequently achieve significant clinical changes.

Furthermore, as the results of our mixed-effects multilevel meta-analysis suggest, this effectiveness might be independent of person-related variables (i.e., gender, age), diagnosis-related variables (i.e., problem behavior topography, diagnosis and nature of the diagnosis) and intervention-related variables (i.e., type of FBA and behavior modification technique used). This could be due to the core characteristic of this kind of intervention: its idiosyncratic adaptation to the contingencies of the person's behavior regardless of its topography, its related diagnosis or its developmental nature. Therefore, as shown by Hurl et al. (2016), considering the functional aspect of a certain problem behavior, regardless of its allegedly bizarre topography, could be a key therapeutic tool to enhance the therapeutic power of our intervention.

However, the absence of a significant difference in the effect size

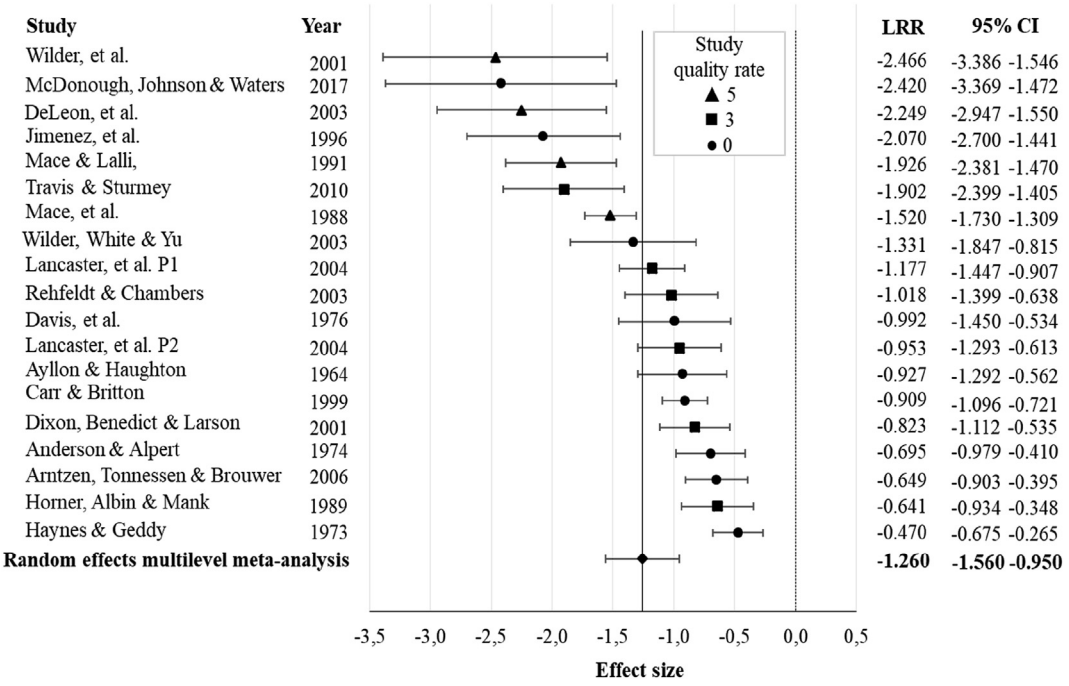

Fig. 2. Results of the random effects multilevel meta-analysis.

Table 5
Summary results of the meta-analysis.

	Studies	Cases	Est.	SE (Est.)	d.f.	95% CI	% reduction	$\hat{\tau}^2$	ω^2
Model 1								0.309	0.001
Overall average	18	19	−1.26	0.142	16.6	[-0.959, −1.56]	[-61.672, −78.986]		
Model 2								0.138	0.002
Publication year	18	19	-.0208	0.0102	7.00	[0.00325, −0.0448]			
QA 0	9	9	−1.1705	0.2216	5.56	[-0.61788, −1.7232]	[-46.091, −82.151]		
QA 3	5	6	−0.9327	0.1454	4.33	[-0.54080, −1.3245]	[-41.772, −73.406]		
QA 5	4	4	−1.9474	0.1501	2.69	[-1.43744, −2.4573]	[-76.246, −91.433]		

across the different FBA methods seems to be somewhat counter-intuitive; it would be expected that interventions based on mixed or experimental FBA methods would be able to detect the environmental controlling variables more precisely and, consequently, modify problem behaviors more effectively (Hurl et al., 2016). This could mean that even indirect functional assessment techniques are precise enough to establish the environmental contingencies of these problem behaviors. However, it could also be due to methodological limitations (e.g., small study sample). This should be further addressed by future research on this topic.

On the contrary, we found significant differences due to both the publication year and the quality analysis index. As abovementioned, the quality analysis index assesses both the fit of the case intervention methodology to the WWC standards and the estimated strength of the evidence of a causal relation (Kratochwill et al., 2010). Given that both the QA-3 and the QA-5 groups indicated moderate evidence of a causal effect and that only QA-5 differed significantly from Q-0, it seems that the actual moderating effect of the quality analysis index lies at its measure of the methodological quality; in other words, case interventions show a larger effect size when the quality of their methodological design is better. This suggests that a good methodological design is needed in order to better appreciate the effectiveness of the intervention. Therefore, we would like to encourage future researchers and practitioners to employ experimental single-case designs in their FBA-based interventions (whenever this is deemed possible according to ethical standards). On the other hand, when combined, the moderating effect of the publication year was at least partially confounded with the moderating effect of the quality analysis index. This suggest that there is a positive time trend towards the enhancement of the methodological quality of the FBA-interventions on atypical vocalizations.

However, this study has obvious methodological limitations that compel us to be cautious when interpreting the results. Firstly, it is not exempt from the characteristic biases of any meta-analysis. Publication bias, for example, could be the reason why only cases with positive results have been included. We have tried to control for this bias by including searches in the gray literature. However, none of the studies found with this resource met our searching criteria. Secondly, although our intervention sample closely approached the 5-3-20 WWC criterion to perform a quantitative synthesis of SCD studies, it still remained small. To account for this potential source of bias, we included small-sample corrections in our meta-analysis. However, our conclusions would be better-informed with a larger study sample with better methodological designs. Thirdly, the amount of residual heterogeneity of our mixed effects multilevel model was still high and significant. Future research should consider other potential moderators of the size effect in order to account for the unexplained variance. Another potential limitation is that this review only included papers with individuals aged 18 and over, it would be very interesting to open up the focus of a future review to adolescents. Finally, it would be very interesting to compare FBA-based interventions with non-FBA-based interventions on atypical vocalizations, in order to determine whether significant therapeutic differences arise.

Despite these limitations, this study is the first to quantitatively synthesize the results of SCD studies of FBA-based interventions for

atypical vocalizations, and we hope that its results will be of help to other mental health practitioners when choosing the appropriate evaluation methods to assess this kind of psychological problems. Overall, our analyses suggest that the FBA might be an effective and efficient method to guide an intervention for the treatment of atypical vocalizations. Furthermore, we believe that this might be interesting to consider in view of the recent debate on the nature and etiology of these phenomena (Deacon & McKay, 2015; Rosernfarb, 2013), since it suggests that behaviors traditionally related to severe and enduring mental illnesses, such as delusions, hallucinations or disorganized speech, could be dependent on environmental contingencies and therefore modifiable through their manipulation. If so, any given intervention on this kind of problems should always take the environmental contingencies into account, regardless of their alleged neurological etiology. The fact that interventions designed from a medical model do not perform a functional assessment of these problems could partially explain why they tend to become chronic. Perhaps, we should start to pay more attention to the functional role of these behaviors in the lives of the people who present them.

Funding

This research was supported by grant PSI 2016-76551-R from the Ministerio de Economía, Industria y Competitividad of the Spanish Government.

Acknowledgements

We would like to thank James E. Pustejovsky for his guidance in the understanding and calculation of the LRR and the subsequent statistical analyses. This research was supported by grant PSI 2016-76551-R from the Ministerio de Economía, Industria y Competitividad of the Spanish Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2019.103444.

References

Alumbaugh, R. V. (1971). Use of behavior modification techniques toward reduction of hallucinatory behavior: A case study. *Psychological Record, 21*(3), 415–417.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA, US: American Psychiatric Publishing, Inc.

Anderson, L. T., & Alpert, M. (1974). Operant analysis of hallucination frequency in a hospitalized schizophrenic. *Journal of Behavior Therapy and Experimental Psychiatry, 5*(1), 13–18. https://doi.org/10.1016/0005-7916(74)90007-X.

APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*(9), 839–851. https://doi.org/10.1037/0003-066X.63.9.839.

Arntzen, E., Tonnessen, I. R., & Brouwer, G. (2006). Reducing aberrant verbal behavior by building a repertoire of rational verbal behavior. *Behavioral Interventions, 21*(3), 177–193. https://doi.org/10.1002/bin.220.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*, 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

Ayllon, T., & Haughton, E. (1964). Modification of symptomatic verbal behaviour of mental patients. *Behaviour Research and Therapy,* 87–97.

Ayllon, T., & Michael, J. (1959). The psychiatric nurse as a behavioral engineer. *Journal of the Experimental Analysis of Behavior, 2*(4), 323–334.

Baruch, D. E., Kanter, J. W., Busch, A. M., & Juskiewicz, K. L. (2009). Enhancing the therapy relationship in acceptance and commitment therapy for psychotic symptoms. *Clinical Case Studies, 8*(3), 241–257. https://doi.org/10.1177/1534650109334818.

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior: Thirty years of functional analysis. *Journal of Applied Behavior Analysis, 46*(1), 1–21. https://doi.org/10.1002/jaba.30.

Burns, C. E., Heiby, E. M., & Tharp, R. G. (1983). A verbal behavior analysis of auditory hallucinations. *The Behavior Analyst. The Behavior Analyst, 6*, 133–143. 6(2, 2), 133 https://doi.org/10.1007/BF03392392.

Carr, E. G. (1977). The motivation of self-injurious behavior: A review of some hypotheses. *Psychological Bulletin, 84*(4), 800–816. https://doi.org/10.1037/0033-2909.84.4.800.

Carr, J. E., & Britton, L. N. (1999). Idiosyncratic effects of noncontingent reinforcement on problematic speech. *Behavioral Interventions, 14*, 37–43.

Common, E. A., Lane, K. L., Pustejovsky, J. E., Johnson, A. H., & Johl, L. E. (2017). Functional assessment–based interventions for students with or at-risk for high-incidence disabilities: Field testing single-case synthesis methods. *Remedial and Special Education, 38*(6), 331–352.

Davis, J. R., Wallace, C. J., Liberman, R. P., & Finch, B. E. (1976). The use of brief isolation to suppress delusional and hallucinatory speech. *Journal of Behavior Therapy and Experimental Psychiatry, 7*(3), 269–275. https://doi.org/10.1016/0005-7916(76)90012-4.

Deacon, B. J., & McKay, D. (2015). The biomedical model of psychological problems: A call for critical dialogue. *The Behavior Therapist, 38*(7), 231–235.

DeLeon, I. G., Arnold, K. L., Rodriguez-Catter, V., & Uy, M. L. (2003). Covariation between bizarre and nonbizarre speech as a function of the content of verbal attention. *Journal of Applied Behavior Analysis, 36*(1), 101–104. https://doi.org/10.1901/jaba.2003.36-101.

Dixon, M. R., Benedict, H., & Larson, T. (2001). Functional analysis and treatment of inappropriate verbal behavior. *Journal of Applied Behavior Analysis, 34*(3), 361–363. https://doi.org/10.1901/jaba.2001.34-361.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* East Norwalk, CT, US: Appleton-Century-Crofts. https://doi.org/10.1037/10627-000.

Haynes, S. N., & Geddy, P. (1973). Suppression of psychotic hallucinations through time-out. *Behavior Therapy, 4*(1), 123–127. https://doi.org/10.1016/S0005-7894(73)80083-8.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., et al. ... Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, d5928. https://doi.org/10.1136/bmj.d5928.

Horner, R. H., Albin, R. W., & Mank, D. M. (1989). Effects of undesirable, competing behaviors on the generalization of adaptive skills. A case study. *Behavior Modification, 13*(1), 74–90. https://doi.org/10.1177/01454455890131005.

Hurl, K., Wightman, J., Haynes, S. N., & Virués-Ortega, J. (2016). Does a pre-intervention functional assessment increase intervention effectiveness? A meta-analysis of within-subject interrupted time-series studies. *Clinical Psychology Review, 47*, 71–84. https://doi.org/10.1016/j.cpr.2016.05.003.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994a). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. https://doi.org/10.1901/jaba.1994.27-197.

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., ... Mazalesk, J. L. (1994b). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis, 27*(2), 215–240. https://doi.org/10.1901/jaba.1994.27-215.

Jimenez, J. M., Todman, M., Pérez, M., Godoy, J. F., & Landon-Jimenez, D. V. (1996). The behavioral treatment of auditory hallucinatory responding of a schizophrenic patient. *Journal of Behavior Therapy and Experimental Psychiatry, 27*(3), 299–310.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*(1), 61–76. https://doi.org/10.1002/jrsm.35.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2010). *Single-case designs technical documentation.* What Works Clearinghouse.

Lancaster, B. M., LeBlanc, L. A., Carr, J. E., Brenske, S., Peet, M. M., & Culver, S. J. (2004). Functional analysis and treatment of the bizarre speech of dually diagnosed adults. *Journal of Applied Behavior Analysis, 37*(3), 395–399. https://doi.org/10.1901/jaba.2004.37-395.

Layng, T. J., & Andronis, P. T. (1984). Toward a functional analysis of delusional speech and hallucinatory behavior. *The Behavior Analyst, 7*(2), 139–156.

Lindsley, O. R. (1956). Operant conditioning methods applied to research in chronic schizophrenia. *Psychiatric Research Reports, 5*, 118–139.

Lutgens, D., Gariepy, G., & Malla, A. (2017). Psychological and psychosocial interventions for negative symptoms in psychosis: Systematic review and meta-analysis. *The British Journal of Psychiatry, 210*(5), 324–332. https://doi.org/10.1192/bjp.bp.116.197103.

Mace, F. C. (1994). The significance and future of functional-analysis methodologies. *Journal of Applied Behavior Analysis, 27*(2), 385–392. https://doi.org/10.1901/jaba.1994.27-385.

Mace, F. C., & Lalli, J. S. (1991). Linking descriptive and experimental analyses in the treatment of bizarre speech. *Journal of Applied Behavior Analysis, 24*(3), 553–562.

Mace, F. C., Lalli, J. S., & Lalli, E. P. (1991). Functional analysis and treatment of aberrant behavior. *Research in Developmental Disabilities, 12*(2), 155–180. https://doi.org/10.1016/0891-4222(91)90004-C.

Mace, F. C., Webb, M. E., Sharkey, R. W., Mattson, D. M., & Rosen, H. S. (1988). Functional analysis and treatment of Bizarre speech. *Journal of Behavior Therapy and Experimental Psychiatry, 19*(4), 289–296. https://doi.org/10.1016/0005-7916(88)90060-2.

Madden, G. J., Dube, W. V., Hackenberg, T. D., Hanley, G. P., & Lattal, K. A. (2013). *APA handbook of behavior analysis, Vol. 1*Washington, DC, US: American Psychological Association. Methods and principles https://doi.org/10.1037/13937-000.

Markwick, L., Smith, C., & Mick, D. (2014). Functional behavioral analysis and social scripting for the older patient with schizophrenia: A staff development program. *Issues in Mental Health Nursing, 35*(11), 883–890. https://doi.org/10.3109/01612840.2014.897777.

McDonough, M. R., Johnson, A. K., & Waters, B. J. (2017). Reducing attention maintained bizarre speech in an adult woman. *Behaviour Analysis: Research and Practice, 17*(2), 197–203. https://doi.org/10.1037/bar0000059.

Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification, 40*(6), 874–900. https://doi.org/10.1177/0145445516645763.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting Items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7) https://doi.org/10.1371/journal.pmed.1000097.

National Institute for Health and Care Excellence (2014). *Psychosis and schizophrenia in adults. Treatment and management, Vol. 178*. NICE Guideline.

Nydegger, R. V. (1972). The elimination of hallucinatory and delusional behavior by verbal conditioning and assertive training: A case study. *Journal of Behavior Therapy and Experimental Psychiatry, 3*(3), 225–227. https://doi.org/10.1016/0005-7916(72)90080-8.

Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological Methods, 20*(3), 342–359. https://doi.org/10.1037/met0000019.

Pustejovsky, J. E. (2018a). clubSandwich: Cluster-Robust (sandwich) variance estimators with small-sample corrections (Version 0.3.2). Retrieved from https://CRAN.R-project.org/package=clubSandwich.

Pustejovsky, J. E.. SingleCaseES: A calculator for single-case effect size indices. HTML. (2018). Retrieved from https://github.com/jepusto/SingleCaseES (Original work published 2016).

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

R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Rehfeldt, 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*(2), 259–261. https://doi.org/10.1901/jaba.2003.36-259.

Rohatgi, A. (2018). WebPlotDigitizer (Version 4.1). Austin, Texas, USA. Retrieved from https://automeris.io/WebPlotDigitizer.

Rosenfarb, I. S. (2013). A functional analysis of schizophrenia. *Psychological Record, 63*(4), 929–946. https://doi.org/10.11133/j.tpr.2013.63.4.013.

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology, 52*(2), 109–122. https://doi.org/10.1016/j.jsp.2013.11.009.

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123–147. https://doi.org/10.1016/j.jsp.2013.11.005.

Skelton, M., Khokhar, W. A., & Thacker, S. P. (2015). Treatments for delusional disorder. Cochrane database of systematic reviews. https://doi.org/10.1002/14651858.CD009785.pub2.

Slade, P. D. (1972). The effects of systematic desensitisation on auditory hallucinations. *Behaviour Research and Therapy, 10*, 85–91.

Sturmey, P., Ward-Horner, J., Marroquin, M., & Doran, E. (2007). Structural and functional approaches to psychopatology and case formulation. *Functional analysis in clinical treatmen* (pp. 1–21). Academic Press.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375–393. https://doi.org/10.1037/met0000011.

Travis, R., & Sturmey, P. (2008). A review of behavioral interventions for psychotic verbal behavior in people with intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities, 1*(1), 19–33.

Travis, R., & Sturmey, P. (2010). Functional analysis and treatment of the delusional statements of a man with multiple disabilities: A four-year follow-up. *Journal of Applied Behavior Analysis, 43*(4), 745–749. https://doi.org/10.1901/jaba.2010.43-745.

Turner, D. T., van der Gaag, M., Karyotaki, E., & Cuijpers, P. (2014). Psychological interventions for psychosis: A meta-analysis of comparative outcome studies. *American Journal of Psychiatry, 171*(5), 523–538. https://doi.org/10.1176/appi.ajp.2013.13081159.

Vandbakk, M., Arntzen, E., Gisnaas, A., Antonsen, V., & Gundhus, T. (2012). Effect of training different classes of verbal behavior to decrease aberrant verbal behavior. *The Analysis of Verbal Behavior, 28*, 137–144.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*https://doi.org/10.18637/jss.v036.i03.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *Lancet (North American Edition), 382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6.

Wilder, D. A., Masuda, A., O'Connor, C., & Baham, M. (2001). Brief functional analysis and treatment of bizarre vocalizations in an adult with schizophrenia. *Journal of Applied Behavior Analysis, 34*(1), 65–68. https://doi.org/10.1901/jaba.2001.34-65.

Wilder, D. A., White, H., & Yu, M. L. (2003). Functional analysis and treatment of bizarre vocalizations exhibited by an adult with schizophrenia: A replication and extension. *Behavioral Interventions, 18*(1), 43–52. https://doi.org/10.1002/bin.128.

Wong, S. E. (2014). A critique of the diagnostic construct schizophrenia. *Research on Social Work Practice, 24*(1), 132–141. https://doi.org/10.1177/1049731513505152.

Zimmerman, K. N., Pustejovsky, J. E., Ledford, J. R., Barton, E. E., Severini, K. E., & Lloyd, B. P. (2018). Single-case synthesis tools II: Comparing quantitative outcome measures. Research in Developmental Disabilities. https://doi.org/10.1016/j.ridd.2018.02.001.