



Cochran's Q Test of Stimulus Overselectivity within the Verbal Repertoire of Children with Autism

Lee Mason^{1,2}  · Maria Otero^{1,3} · Alonzo Andrews^{4,5}

Accepted: 26 August 2021 / Published online: 15 September 2021

© Association for Behavior Analysis International 2021

Abstract

Stimulus overselectivity remains an ill-defined concept within behavior analysis, because it can be difficult to distinguish truly restrictive stimulus control from random variation. Quantitative models of bias are useful, though perhaps limited in application. Over the last 50 years, research on stimulus overselectivity has developed a pattern of assessment and intervention repeatedly marred by methodological flaws. Here we argue that a molecular view of overselectivity, under which restricted stimulus control has heretofore been examined, is fundamentally insufficient for analyzing this phenomenon. Instead, we propose the use of the term “overselectivity” to define temporally extended patterns of restrictive stimulus control that have resulted in disproportionate populations of responding that cannot be attributed to chance alone, and highlight examples of overselectivity within the verbal behavior of children with autism spectrum disorder. Viewed as such, stimulus overselectivity lends itself to direct observation and measurement through the statistical analysis of single-subject data. In particular, we demonstrate the use of the Cochran Q test as a means of precisely quantifying stimulus overselectivity. We provide a tutorial on calculation, a model for interpretation, and a discussion of the implications for the use of Cochran's Q by clinicians and researchers.

Keywords Autism spectrum disorder · Stimulus overselectivity · Restricted stimulus control · Cochran's Q

Half a century ago, Lovaas et al. (1971) used the term “stimulus overselectivity” to refer to the restricted stimulus control over the behavior of individuals with autism spectrum disorder (ASD) in the presence of complex stimuli. In particular, they defined overselectivity as a “problem of dealing with stimuli in context, a problem of quantity rather than quality of stimulus control” (p. 219). That is,

✉ Lee Mason
lee.mason@cookchildrens.org

despite the availability of multiple stimuli within the environment to signal the availability of reinforcement, a circumscribed repertoire develops under a paucity of reinforcers (Lovaas et al., 1979).

Over the last 50 years, multiple examples of stimulus overselectivity have appeared within the literature. Across these studies, various patterns can be abstracted that together encompass a definition of restricted stimulus control. From the outset, Lovaas et al. (1971) found that even typically developing children will sometimes display selective responding, and that some children with ASD may not display selective responding, therefore highlighting the need for a control group when assessing stimulus overselectivity in individuals with ASD. However, this tends to be the exception rather than the rule (Lovaas et al., 1979).

Researchers have recently begun to question the extent to which participants with ASD demonstrate disproportionate rates of stimulus overselectivity when compared to typically developing peers and peers with other intellectual disabilities. For instance, using a delayed match-to-sample procedure across 37 neurotypical children, Reed et al. (2013) found that typically developing children “cannot be expected to reliably respond to simultaneous multiple cues until after 3 years of age” (p. 1254). The authors concluded that overselectivity is part of general childhood development and not specific to individuals with ASD.

Using a similar visual matching procedure, Rieth et al. (2015) assessed 41 children with ASD for the presence of restricted stimulus control. Only 20% of participants demonstrated overselective responding, leading the authors to conclude that the population of individuals with ASD has changed over time. Overselectivity as described by Lovaas et al. (1971) may not be as prevalent as once thought. Rieth et al. recommend systematically testing for the presence or absence of overselectivity to determine whether incorporation of conditional discriminations with compound stimuli into teaching is necessary for that individual.

Finally, Dube et al. (2016) employed a series of multiple delayed matching procedures to assess the prevalence of overselectivity across three populations: individuals with ASD, individuals with Down syndrome, and typically developing peers. Although they found no significant difference across groups of participants, the authors observed, “. . . overselectivity may occur with some types of stimuli and not others, depending on the relationships among stimulus complexity, learning history, and developmental level” (p. 233).

Contrary to much of the early work on stimulus overselectivity, contemporary researchers have concluded that issues of restricted stimulus control are not specific to individuals with ASD. However, a closer examination of the methodologies that led to this conclusion sheds light on the discrepancy between early and recent findings. At first, Lovaas et al. (1971) employed a multidimensional stimulus display with complex, incongruous stimuli. The authors acknowledged, “Our data failed to support notions that any one sense modality is impaired in autistic children, or that a particular sense modality is the ‘preferred’ modality” (p. 219). However, rather than examining a wider variety of stimuli and exteroceptive modalities, the ensuing trajectory of research on stimulus overselectivity has ironically become restricted to methodologies of visual matching.

Underestimating Overselectivity

Gamba et al. (2015) noted a prolific lack of construct validity within the assessment of verbal behavior, and contemporary approaches to assessing stimulus overselectivity appear to suffer the same limitation. That is, the measures used in these and other studies of overselectivity may not actually assess what they purport. A commonality across studies of overselectivity is the appearance of a training phase in which the investigators condition participants to respond to a multicomponent stimulus so that they can later assess the response strength of each component individually. Although this approach appears to have had some success in identifying the restricted controlling relations, the continual use of a match-to-sample operandum is contrary to Lovaas's finding that no particular sense modality (i.e., visual matching) is preferred among individuals with ASD. Given that matching-to-sample is not a core deficit of ASD, there should be no expectation of substantial differences in the prevalence of overselectivity when compared to neurotypical peers (cf. Dube et al., 2016; Reed et al., 2013; Rieth et al., 2015). The exclusive use of visual matching to assess overselectivity hinders our understanding of this phenomenon.

Rieth et al. (2015) argued that the population of individuals with ASD had changed due to the advancement of intervention procedures designed to reduce overselectivity, but failed to recognize that the assessment procedures used to identify stimulus overselectivity have become increasingly narrow in scope. In summarizing the literature on overselectivity, Brown and Bebko (2012) stated, “. . . within and between each modality, there does not seem to be a constant local aspect of the stimuli towards which the majority of children with autism preferentially attend, making it difficult to predict which aspect of stimuli would be associated with a particular trained response” (p. 735). Given the wide range of exteroceptive modalities and stimuli classes that overselectivity may encompass, the premise that such a pervasive disorder can be captured within a simple match-to-sample framework appears to be invalid for examining the prevalence of overselectivity across populations for three primary reasons. First, the modalities assessed may not be sufficiently broad in scope to observe restrictive stimulus control. Next, the duration of training may not be long enough to establish restricted control. Finally, the participants most likely to exhibit overselectivity may have been excluded.

Lovaas et al. (1971) cited participant selection as the most important qualification for researching overselectivity, and Rieth et al. (2015) recommend developing an assessment tool to identify participants who exhibit overselectivity across different exteroceptive modalities. The singular focus on identifying overselectivity through matching-to-sample may only be reasonable for participants who have a history of visual fixation or ocular stereotypy (Kelly & Reed, 2020). Ergo, a more pragmatic starting place for researching overselectivity within the population of individuals with ASD is with skill deficits consistent with the diagnostic criteria and functional dimensions of the disorder: communication and social interaction (American Psychiatric Association, 2013; Ibrahim & Sukhodolsky, 2018).

The extant research on stimulus overselectivity has another common flaw in that for overselectivity to be assessed, a novel response must first be taught. It is notable that a half century of studies attempting to assess overselectivity all required a training component. The unspoken assumption here is that the novel apparatus will

simultaneously exert both generalized *and* restricted stimulus control over the participant's responding, using Stokes and Baer's (1977) least preferred method of train and hope. Although the training sessions are long enough to condition the assessment task, they may not be long enough to condition restricted stimulus control. Researchers in this area have assumed that stimulus control is restricted from the outset, but this is not necessarily the case. Additional research is needed to clarify whether restricted stimulus control would develop over additional training trials.

Finally, the results of participants who fail to acquire the training task are often excluded from analysis. Might these participants be the ones mostly like to exhibit overselectivity? In addition to the narrow scope of contemporary research on stimulus overselectivity failing to capture the nature of this phenomenon, they also fail to produce outcomes that are socially valid for use in clinical applications or aide in the amelioration of overselectivity.

A modification of Wing's (1988) triad model conceptualizes ASD as specific deficits of communication, reciprocity, and behavior excesses associated with repetitive behavior and circumscribed interests. Visual acuity is not part of the diagnostic criteria for ASD, yet this is the modality on which the preponderance of research on overselectivity has focused. Gersten (1980) called for research on overselectivity to be imbued with more socially valid outcomes, such as language acquisition and concept formation.

In terms of social significance, Ferster et al. (1975) described verbal behavior as "the ultimate application of principles of stimulus control to human behavior" (p. 551). Although additional research is necessary to bridge the divide between overselectivity in visual performance and overselectivity in verbal behavior, there is no reason to suspect that one begets the other. Inasmuch as children with ASD have challenges interacting and communicating, their language deficits may indicate stimulus overselectivity. Some children with ASD show global language deficits, whereas others show an extensive vocabulary under circumscribed control. Through early intensive behavioral intervention, verbal behavior can be established and stimulus control expanded to a more comprehensive range; that is, from the self to the listener-other. As Baron-Cohen (2005) points out, "[t]he term 'autism' literally means 'self'-ism, derived from the Greek word 'auto' ('self')" (p. 166), and as such autism represents a disorder of being overly self-focused. Verbal behavior effectively becomes the bridge from self towards others.

Understanding Overselectivity

In summarizing the behavior-analytic literature on stimulus overselectivity, Ploog (2010) observed that it "falls short for investigating a phenomenon that is too narrowly conceptualized and that heavily depends on a given experimental paradigm" (p. 1334). Rather than attempting to assess stimulus overselectivity within the confined scope of an arbitrarily selected and predefined task, we posit that meaningful differences in stimulus control can only be observed through temporally extended activities that document a pervasive history of selective responding. The ontogenetic conditioning of disproportionate levels of stimulus control is apparent across the verbal repertoire of individuals with ASD. Lovaas et al. (1971) argued, "... speech exists without meaning

to the extent it has an impoverished context” (p. 221). The verbal operants identified by Skinner (1957) provide the context through which we can examine disproportionate stimulus control.

Foreshadowing the outcomes of future research on overselectivity, Lovaas et al. (1971) stated that, “[p]erhaps the most important qualification centers on the choice of [participants] and the bases of their diagnoses. It is noteworthy that we have worked with the most regressed of autistic children, and that different results may have been obtained had we used children who were more advanced, having, for example, speech development” (p. 219). In contrast, in each of the studies that found no significant differences in overselectivity between individuals with and without ASD, the participants appear to have had advanced listening and speaking repertoires.

The differences in participants’ language skills between the findings of early and contemporary research on overselectivity is no mere coincidence. Although language skills are not necessarily indicative of restricted stimulus control (Ploog, 2010), overselectivity may be better assessed by more closely examining the verbal repertoire. As we have noted, overselectivity cannot be responsible for verbal behavior deficits because extended temporal patterns of disproportionate stimulus control over molar activities like language are precisely what we call overselectivity. Contextualizing the verbal operants as different populations of responding affords the use of nonparametric statistical procedures to help differentiate prepotent control from the noise of random variation (Davison, 1999).

Populations of Performance

Skinner (1973) explained reinforcement as an increase in the probability of a given response or class of responses. According to Skinner,

As our knowledge of the effects of contingencies of reinforcement increases, we can more often predict what an organism will do by observing the contingencies; and by arranging contingencies, we can increase the probability that an organism will behave in a given way. In the latter case, we may be said to “control” its behavior. The term does not mean forcible coercion or the triggering of a jack-in-the-box kind of reflex action. . . . Human behavior is controlled not by physical manipulation but by changing the environmental conditions of which it is a function. The control is probabilistic. The organism is not forced to behave in a given way; it is simply made more likely to do so. (p. 259)

As a result, we can contextualize the notion of molecular behavior as instances of an operant class from which the emission of a given topography can be estimated. Skinner’s discussion of response probabilities under the control of certain environmental conditions gives way to a statistical analysis of operant behavior, which directly relates to our discussion of stimulus overselectivity.

Guttman and Kalish (1956) describe the relationship between discriminative control and generative control as a continuum upon which operant behavior is more or less probable. This continuum provides the basis for a probability density function in which the relative likelihood of behavior is greatest at its centrality and weakest at either tail.

For instance, having been conditioned to respond “ball” in the presence of baseballs and basketballs, untrained exemplars from the operant class that share common features of the trained stimuli (e.g., volleyballs and tennis balls) will be more likely to exert control over this response, relative to both members of the class that do not share these controlling features (e.g., rugby balls and footballs) and nonmembers of the class that do share these controlling features (e.g., rocks and scoops of ice cream). Over time, a history of reinforcement yields a distribution of environmental control over each operant class of behavior.

Further research on generalization gradients has shown that environmental control is normally distributed (Honig & Urcuioli, 1981), and individuals whose behavior follows this model of stimulus control are frequently referred to as “typically developing,” “neurotypical,” or—in accordance with their behavioral distribution—“normal.” In contrast, individuals whose behavior consistently skews from this model in particular ways may receive other labels to describe atypical stimulus control. For example, individuals who draw metaphorical extensions between otherwise unrelated events might be regarded as “artists” or “poets.” Likewise, individuals who characteristically respond to only a selective range of the environment may receive a diagnosis of ASD.

Although the normal distribution may be convenient for conceptualizing the probability of a given response within an operant class, it does not fully account for the greater environmental realm in which multiple controlling variables compete with one another. For instance, in the above example, we categorized the response “ball” in the presence of a rock as an error of commission. However, given the requisite history, it would be just as valid to categorize faulty stimulus control over the response “rock” as an error of omission. Consistent with Davison’s (1999) argument, “. . . high levels of experimental control lead inevitably to the nonnormality of data distributions” (p. 102), we posit that a one-sided distribution, such as X^2 , provides a more representative model for studying stimulus overselectivity within the applied setting. (See Bakker & Wicherts, 2011, and Fairbanks & Rytting, 2001, for a comprehensive discussion of X^2 .)

A renewed understanding of stimulus overselectivity gives way to new avenues for the observation, measurement, and study of restricted stimulus control. The purpose of the current article is to expand upon the quantitative analysis of individual behavior by introducing the Cochran Q test as a measure of relative differences in stimulus control over the verbal behavior of individuals with ASD. Behavior analysts have previously argued for the use of nonparametric statistical procedures to formalize the process of assessment and serve as discriminative stimuli for interpreting data (Davison, 1999). In particular, a significant p -value allows the reader to distinguish real, reliable effects from random variation in independent variables. Based on the X^2 distribution, Cochran’s Q is of particular interest to clinicians who work with individuals with ASD due to its ability to tell us whether the individual’s behavior shows a stimulus control deficit (i.e., statistically significant at the .05 level, α), along with the magnitude of that deficit through its effect size, R .

Cochran Q Test

Cochran (1950) introduced Q as a comparison of percentages in matched samples for more than two dichotomous dependent variables. Cochran sought to determine whether

binary distributions represented similar (i.e., null hypothesis) or different (i.e., alternative hypothesis) distributions, but was unable to do so using conventional analyses. Said Cochran (1950), “a mixture of 1’s and 0’s could not by any stretch of the imagination be regarded as normally distributed” (p. 262). Unlike a one-way, repeated-measures ANOVA, Cochran’s Q is based on the X^2 distribution. It is frequently used in meta-analyses to assess the heterogeneity of effect-size estimates from individual studies (Hoaglin, 2016). Cochran’s Q is analogous to a Freedman test for use with a binary dependent variable (e.g., correct/incorrect, response/no response). Thus, Cochran’s Q lends itself to the quantitative analysis of discrete trial data.

Cochran’s Q is considered a generalization of the McNemar test, with Q allowing for greater degrees of freedom. McNemar’s test examines the relationship between two dependent samples, like quantitative models of behavioral choice and signal detection. Quantitative analyses of behavior have also provided sensitive measures of overselectivity. However, their extension to verbal behavior is scant

Cochran’s Q may have utility for studying the verbal behavior of individuals with ASD, by providing a novel analysis for examining a dichotomous dependent variable (i.e., the presence or absence of a given response topography) across various environmental conditions (i.e., tact, mand, and two types of intraverbal control). A binary notation of 1 or 0 denotes whether or not the target response occurred under each stimulus condition. For example, the response “rock” may occur under tact (1) and mand (1) but not intraverbal control (0), whereas “ball” may occur under tact (1) and intraverbal (1) control but show no strength as a mand (0). After repeating this exercise across multiple targets, the total number of responses (1’s) are then summed for each verbal operant class. “The problem,” wrote Cochran, “is to test whether these totals differ significantly” (p. 258). Null hypothesis significance testing may be useful to distinguish between random variability and stimulus overselectivity.

Operant Units of Analysis

Statistical analyses based on group designs typically regard the individual participant as the unit of analysis. When analyzing response populations within an organism, however, related samples are those that share formal similarity. A primary assumption is that responses with similar topographies also share similar strength across operant classes. For instance, when we say that a person knows the word “ball,” we imply that they can just as easily emit the response when labeling the presence of a ball, as when requesting access to a ball. Although both instances of the response are functionally distinct, they are topographically similar, and we may therefore assume that success and standard error rates are equivalent.

Skinner’s (1957) *Verbal Behavior* identified the manner in which verbal operants function independently of one another. When analyzing verbal behavior, the use of a related dependent variable means examining the same verbal response under various environmental conditions: tact, mand, echoic, and sequelic. Prior research has shown that the functional language of individuals with ASD often develops at different rates (Sundberg, 2007). For example, it is not uncommon to meet a speaker with autism whose verbal behavior is overly selective of tact or echoic sources relative to mand or sequelic control. The child may be able to label an item when it is present, but cry to access it once it has been removed. They may echo the name of an item, but fail to

respond when provided with a description of it. Occasional errors or omissions are common in everyday speech, and should likewise be expected of speakers with ASD. However, behavior analysts seeking to explain, predict, and control the extent to which a speaker is likely to emit verbal behavior under various sources of control must be able to distinguish current environmental factors from histories of conditioning. As a more comprehensive way of measuring stimulus overselectivity, we are interested in determining whether disproportionate rates of responding under each of four environmental conditions are statistically significant. Cochran's Q allows for a comparison of response strength across operant classes by systematically testing verbal behavior under different sources of control.

Identifying Overselectivity with Cochran's Q Test

Here we extend Lerman et al.'s (2005) procedures for functionally analyzing the verbal behavior of children with developmental disabilities to meet the requirements of Cochran's Q test. The procedures described by Lerman et al. have been previously extended to include a broader range of participants (LaFrance et al., 2009) and simplify the use of controls (Normand et al., 2008). Kelley et al. (2007) further refined these procedures to control for the number of trials across conditions.

A verbal operant experimental analysis systematically alters multiple environmental conditions to assess the occurrence or nonoccurrence of specific verbal responses. Ensuring that a listener was present to reinforce all responses, Lerman et al. (2005) assessed the controlling variables for individual words frequently spoken by each participant that did not consistently operate on the environment in any particular way. The target word(s) for each individual were assessed across four isolated sources of strength. In assessing the presence of mand control, access to the target object was restricted, while carefully controlling for tact, echoic, and sequelic variables. In contrast, tact control was assessed by presenting the target object to the speaker, while ensuring the absence of confounding mand, echoic, and sequelic sources. For assessing echoic control, the name of the target object was spoken, whereas sources of tact, mand, and sequelic control were removed. Finally, sequelic control was assessed by developing an intraverbal fill-in for each target response, while eliminating sources of tact, mand, and echoic control.

The conditions described by Lerman et al. (2005) for isolating the sources of control across four verbal operants are directly transferable to our purposes in completing a Cochran's Q analysis. Each source of control over the speaker's verbal behavior is experimentally manipulated while eliminating confounding factors. However, whereas Lerman et al. sampled the language of their participants with one or two responses, the Cochran's Q test requires a more robust sample size to determine the existence of a statistically significant difference.

Table 1 shows an example of how a verbal operant experimental analysis can be formalized for nonparametric statistical analysis (Davison, 1999). The participant selects three different items to start each session. The names of these three items are then assessed under independent sources of tact, mand, echoic, and sequelic control. This process is repeated up to four times, using different items for each session to assess a variety of verbal responses.

Table 1 Sequencing of conditions across multiple sessions of a verbal operant experimental analysis

Order of Conditions	Session 1: Responses 1–3	Session 2: Responses 4–6	Session 3: Responses 7–9	Session 4: Responses 10–12
First	Tact	Tact	Tact	Tact
Second	Mand	Echoic	Mand	Sequelic
Third	Echoic	Sequelic	Sequelic	Mand
Fourth	Sequelic	Mand	Echoic	Echoic

Note. The same three response topographies are assessed across all four conditions of each session. Two to four sessions are required for Cochran’s *Q* test

Equating the number of response opportunities across conditions allows for a direct comparison of the different controlling properties. It should be noted that rich schedules of reinforcement may influence response bias. The verbal operant experimental analysis controls for this by using a topography-based dependent variable, and by reinforcing each response topography a maximum of once in each condition. As in traditional functional analyses of challenging behavior, the continuous alteration of conditions across multiple sessions also mitigates carryover and sequencing effects. However, we cannot entirely rule out differences in reinforcer variables across conditions (Hannula et al., 2020).

The Power of *N* = 1

In summarizing various conceptual frameworks for analyzing behavior, Moxley (1987) concluded, “The choice of a unit, however, does make a difference in our interpretation and responsiveness to those events” (p. 24). Much as the three-term contingency provides the context for analyzing behavior, Cochran’s *Q* test is premised upon an adequate unit of analysis. Adequate power is also critical in *N*-of-1 research, as the effect-size ranges for single-subject data are typically much larger than those proposed for group-design studies (Dowdy et al., 2021; Kyonka, 2019). Mason and Andrews (2019) previously proposed that 30 responses were required to sufficiently power a verbal operant experimental analysis. To this, we echo Tate and Brown’s (1970) response to McNemar’s (1962) declaration that *Q* only follows the X^2 distribution when *n* is large: “His rule of thumb, $N > 30$, is unrealistic and inconsistent with Cochran’s check on the accuracy of *Q* in small samples” (p. 156). A simple power analysis allows us to gather a sufficient sample to minimize chance findings for a given effect size.

Underpowered studies (i.e., too few samples) may have medium to large effects without yielding statistically significant *p*-values. For instance, the emergent speakers in Hall and Sundberg’s (1987) study demonstrated strong tact control in comparison to low levels of mand function. Although the effect was large, the statistical significance between tact and mand control can only be determined with a larger sample size.¹

¹ Hall and Sundberg (1987) assessed tact control over 10–11 items, but only assessed mand control over 4 of those items. Because only four responses (*r*) were assessed across both conditions (*c*), the sample size (*r***c*) of *n* = 8 is insufficient for statistical analysis.

Similar research on generalization gradients have found that low response rates artifactually produce steep gradients (Honig & Urcuioli, 1981).

In contrast, overpowered studies (i.e., too many samples) may show statistical significance, but provide inconsequential effects. For example, Davis's (2017) analysis of high-school students' verbal behavior found an overwhelming preponderance of intraverbal control. Despite any statistical significance, the vast sample size (approximately 24 hr of audio/video recordings) nullified any meaningful effect.

Adequately powered studies have the ability to determine statistical significance in conjunction with moderate to large effects. They are also particularly useful in determining the extent to which Cochran's Q approximates the X^2 distribution. Tate and Brown (1970) extensively analyzed the distributions of Q in small samples to power the Cochran Q test sufficiently. To assist with practical interpretation, the authors developed a judgmental aid for determining when the X^2 approximation is satisfactory:

Judging from the distributions examined, the chi-squared approximation to Q seems good enough, on the average, for practical work with samples yielding tables of 24 or more scores. The following rule of thumb is suggested for use of the approximation: Delete rows containing only 1's or only 0's. If the product of the number of remaining rows, r , times the number of columns is 24 or more, the approximation is generally satisfactory, provided r is at least 4. (p. 159)

Assessing verbal behavior across four conditions—tact, mand, echoic, and sequelic—requires a minimum sample size of six unique responses: precisely the number of subjects recommended by Davison (1999) for the use of non-parametric statistical analyses. The emission of six responses (r) analyzed across four conditions (c) yields the minimum sample size of 24 items for analysis.

Table 2 shows the results of a verbal operant experimental analysis conducted on a 4-year-old boy with ASD. The tabular data clearly show that some operants (i.e., tact and echoic) were more likely to control the speaker's verbal behavior than others (i.e., mand and sequelic). However, the extent to which these differences in response strength are a simple result of sampling—whether they represent different distributions—remains unclear.

In calculating the significance of differences in Table 2, the null hypothesis tells us that the variation in strength across verbal operants is simply a result of random sampling, and that there is no real variation in these scores. In other words, the population proportions are equal.

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4$$

In contrast, the alternative hypothesis is that the differences in frequencies across the four operants are disproportionate, indicating the presence of stimulus overselectivity.

$$H_A : \neq H_0$$

Table 2 Results of a functional analysis on the verbal behavior of a 4-year-old boy with ASD

Responses	Tact	Mand	Echoic	Sequelic	Total (u_i)
iPad	0	0	1	0	1
Guitar	1	0	1	0	2
Spider	1	0	1	1	3
Lion	1	1	1	0	3
Phone	1	0	0	0	1
Dot	1	1	1	0	3
Total (T_j)	5	2	5	1	$\sum T_j=13 = \sum u_i$

Note. Archival data. 1 = the response at left was emitted for the corresponding condition; 0 = the response at left was not emitted for the corresponding condition. The assessment was conducted in English, the primary language spoken in the child’s home

Cochran’s Q test for differences between the percentage (or proportion) of related samples allows us to assess the significance of these differences. As written by Cochran (1950):

$$Q = \frac{c(c-1) \sum_j (T_j - \bar{T})^2}{c \left(\sum_i u_i \right) - \left(\sum_i u_i^2 \right)} \tag{1}$$

where c is the number of conditions (i.e., columns), T_j is the frequency of successes in the j th column, \bar{T} is the mean of T_j , and u_i is the frequency of successes in the i th row. Tate and Brown (1970) provide an equivalent formula for Q that is more amenable to computation:

$$Q = \frac{(c-1) \left[c \sum_j T_j^2 - \left(\sum_j T_j \right)^2 \right]}{c \sum_i u_i - \sum_i u_i^2} \tag{2}$$

Using the revised formula, we can calculate Q using the data from Table 2, where the problem is to determine whether the four operants vary significantly in strength. Given that $c = 4$, consider the row and column totals, so that

$$\sum_j T_j^2 = 5^2 + 2^2 + 5^2 + 1^2 = 55$$

and

$$\sum_i u_i^2 = 1^2 + 2^2 + 3^2 + 3^2 + 1^2 + 3^2 = 33$$

We can now plug in the variables to complete our analysis using a simple calculator:

$$Q = \frac{3 \left[4(55) - (13)^2 \right]}{4(13) - 33} = 8.053 \quad (3)$$

A summary of the analysis of Table 2 with the Cochran Q test is as follows: Cochran's Q test found a significant difference in stimulus control over the participant's verbal behavior. A statistical table of the X^2 distribution tells us that the critical value for three degrees of freedom with an alpha of .05 is 7.81 (Cohen, 2013). The result can therefore be summarized as $Q(3) = 8.05, p < .05$.

Given that $Q = 8.05$ is larger than the critical value of 7.81, we can interpret this as a pattern of overselectivity (i.e., $p < .05$). In the above example, the value for Q was only slightly greater than the critical value. These data can be contrasted with the dataset in Table 3, which shows the results of a reassessment of the child's functional language skills after five months of behavior-analytic intervention.

Using the same procedures described above, we calculate a Q of 5.00, which is below the critical value for an alpha of .05. Five months of behavior-analytic intervention has reduced the speaker's Q value by 3.05, and—despite the discrepancy in operant strength within the speaker's verbal repertoire—the speaker no longer shows a pattern of stimulus overselectivity (i.e., $p > .05$). The just-noticeable difference between Tables 2 and 3 may not appear to be clinically significant when examining the raw data (Kyonka, 2019). However, the effects of intervention can be seen through the quantitative treatment of single-subject data.

Exclusion Criteria

An important consideration in sampling the verbal repertoire is that rows containing only 1's or 0's have no effect on the value of Q , and therefore must be omitted from our sample. Therefore, speakers with severely limited verbal repertoires are likely to be exempt from Cochran's Q analysis, as they will have a high frequency of 0's across all four verbal operants. Likewise, near fluent speakers are likely to be excluded due to a

Table 3 Results of a reassessment of the verbal behavior of a 4-year-old boy with ASD

Responses	Tact	Mand	Echoic	Sequelic	Total (u_i)
Puzzle	1	0	1	0	2
Cheez-It	1	1	0	0	2
Water	1	0	0	1	2
Oreo	1	1	1	0	3
Bulldozer	0	0	1	0	1
Shark	1	1	1	0	3
Total (T_j)	5	3	4	1	13

Note. Archival data. 1 = the response at left was emitted for the corresponding condition; 0 = the response at left was not emitted for the corresponding condition

preponderance of 1's. Table 4 shows the data collected on the verbal behavior of a 3-year-old girl with ASD. Note that the fourth and fifth stimuli were excluded because this response occurred across all four conditions. Responses that failed to occur across any conditions would be similarly excluded.

In consequence, and in alignment with the guidance of Tate and Brown (1970), our analysis was sufficiently powered such that we can confidently use the critical values of χ^2 to determine the significance of our results from Table 4, $Q = \chi^2 = 12.50$. A summary of the analysis of Table 4 with the Cochran Q test is as follows: Cochran's Q test found a significant difference in stimulus control over the participant's verbal behavior, with $Q(3) = 12.50, p < .05$.

Table 5 shows another assessment in which the data for the first stimulus were eliminated from the analysis because the speaker failed to emit the response across any condition. Additional data were collected to ensure a sufficient sample size for Cochran's Q test.

The prepotency of echoic control is apparent prior to calculating Cochran's Q . This speaker was more likely to emit a response when provided with a verbal imitative stimulus, when compared to other conditions. The Cochran Q test can verify that the restrictive pattern of responding cannot be attributed to random chance or momentary interference, and it provides a precise measure of stimulus overselectivity: $Q(3) = 21.48, p < .001$.

Quantifying the Magnitude of Overselectivity

The mere fact that any observed differences in population distributions are statistically significant conveys little of practical importance about the variables under investigation (Young, 2018). In addition to the usual probability values, applied scientists will find a measure of effect size somewhat more pragmatic. Effect sizes measure the magnitude of a phenomenon, and—when multiplied by 100—may be reported as a percentage.

Berry et al. (2007) suggest a chance-corrected effect size, R , among the response targets (r) over the conditions (c). Though somewhat more complicated to calculate, R is entirely data dependent, and corrected for chance. Therefore, given the null

Table 4 Results of a functional analysis of the verbal behavior of a 3-year-old girl with ASD

Responses	Tact	Mand	Echoic	Sequelic	Total (u_i)
<i>Mamá</i>	1	0	1	0	2
<i>Bebé</i>	1	0	1	0	2
<i>Cun</i>	1	0	1	1	3
<i>Agua</i>	1	1	1	0	3
<i>Autobus</i>	1	0	1	0	2
<i>Perro</i>	1	0	1	1	3
<i>Poppy</i>	1	1	1	0	3
Total (T_j)	7	2	7	2	18

Note. Archival data. 1 = the response at left was emitted for the corresponding condition; 0 = the response at left was not emitted for the corresponding condition. The assessment was conducted in Spanish, the primary language spoken in the child's home

Table 5 Results of a functional analysis of the verbal behavior of a 4-year-old girl with ASD

Responses	Tact	Mand	Echoic	Sequelic	Total (u_i)
Ball	0	0	1	1	2
Apple Jacks	0	0	1	0	1
Rope	0	0	1	0	1
Shoes	0	0	1	0	1
Popcorn	0	0	1	0	1
Horse	0	0	1	0	1
Token	0	0	1	0	1
Baby	0	0	1	0	1
Total (T_j)	0	0	8	1	9

Note. Archival data. 1 = the response at left was emitted for the corresponding condition; 0 = the response at left was not emitted for the corresponding condition

hypothesis of Cochran’s Q , which states that the distribution of δ assigns equal probability to each of c distributions, the average value of δ can be written as

$$\mu_\delta = \frac{2}{c(c-1)} [(\sum_{i=1}^r \rho_i)(r - \sum_{i=1}^r \rho_i) - \sum_{i=1}^r \rho_i(1 - \rho_i)] \tag{4}$$

and the chance-corrected measure of effect size can then be calculated as

$$R = \frac{1}{4} \left(1 - \frac{\delta}{\mu_\delta} \right) \tag{5}$$

Similar to Pearson’s product-moment correlation coefficient, R is zero under chance conditions, one when agreement among response targets is perfect, and R will be negative under conditions of disagreement. As for interpreting the size of effect, R may be interpreted similar to other effect sizes for associations among categorical variables, where $< .10$ is considered negligible, small is $.10 < .30$, medium is $.30 < .50$, and $\geq .50$ is large.

Note. *small, **moderate. † $p < .05$, †† $p < .001$. ‡small, ††medium, †††large

Table 6 shows the quantitative values of verbal behavior for each of the datasets presented above to provide a more direct comparison of the different outcomes. These metrics include the stimulus control ratio equation (SCoRE; Mason & Andrews, 2019), Cochran’s Q test, and chance-corrected effect size, R . When initially assessed, the verbal repertoire of the speaker in Table 2 demonstrated a moderate verbal behavior SCoRE (.58) that showed a small degree (26%) of stimulus overselectivity. When reassessed (Table 3), his SCoRE had increased by .10, and his verbal behavior no longer indicated a pattern of restrictive stimulus control. The speaker whose data appear in Table 4 demonstrated a moderate SCoRE (.69) slightly larger than the speaker in

Table 6 A comparison of verbal behavior metrics

Speaker Data	SCoRE	Cochran’s <i>Q</i>	R
Table 2	.58**	8.05†	.26‡
Table 3	.68**	5.00	.12‡
Table 4	.69**	16.55††	.33‡‡
Table 5	.22*	21.48††	.71‡‡‡

Table 3 (.68), but her data also indicated a medium degree of overselectivity (33%). Finally, the speaker in Table 5 had a small verbal repertoire (SCoRE = .22), and showed that her verbal behavior was largely restricted (71%). The size of the speaking repertoire indicated by the verbal behavior SCoRE is not necessarily indicative of stimulus overselectivity. However, restricted stimulus control may constrain the development of the speaker’s verbal repertoire.

Upon Further Analysis

Post-hoc analyses (e.g., Dunn, 1964) can pinpoint specific relationships when Cochran’s *Q* is significant. For example, Table 7 shows the results of a post-hoc analysis for the data from Table 4, which had a significant main effect of $Q(3) = 16.55, p < .001, R = .33$. After a Bonferroni adjustment of the *p*-value, four relationships were found to be significant.

Inferential statistics produces clear discriminative stimuli for data-based decision making (Davison, 1999). In Table 7, the post-hoc analysis that accompanies the Cochran *Q* test allows researchers and clinicians to more readily discriminate real differences (i.e., the Tact–Mand, Tact–Sequelic, Echoic–Mand, and Echoic–Sequelic relationships) from those that are no greater than chance (i.e., Tact–Echoic and Sequelic–Mand). Cochran’s *Q* test shows a moderate degree of stimulus overselectivity within the speaker’s verbal repertoire, with a prepotency for both tact and echoic sources over sequelic and mand variables. There was no difference between tact and

Table 7 Dunn’s Post-Hoc analyses with a bonferroni correction for multiple comparisons

Comparisons	Sig.	Adj. Sig.
Tact-Mand	.00**	.01*
Tact-Echoic	1.00	1.00
Tact-Sequelic	.01**	.05*
Echoic-Mand	.00**	.01*
Sequelic-Mand	.66	1.00
Echoic-Sequelic	.01**	.05*

Note. **p* < .05, ***p* < .01. For each comparison, the stronger operant is listed before the weaker operant. Significance was adjusted by multiplying the uncorrected *p*-value by the number of comparisons. The maximum *p*-value cannot exceed 1.00

echoic sources of control, and the relationship between sequelic and mand sources, although discrepant, was not found to be significant. In other words, the speaker is more likely to emit a response when the item is present or when given an imitative verbal stimulus, than when access to the item has been restricted or when given a statement about the item.

Aside from offering an unparalleled level of precision for assessing the verbal repertoire, the results of Cochran's Q test have clinical implications for individualizing behavior-analytic intervention. For example, statistically significant differences within the verbal repertoire signify the need for behavior-analytic intervention. In addition, speakers with larger Q values may require services that are more intensive than those with lower Q values. An important caveat to this claim is that individuals with little-to-no functional verbal repertoire will also show low values for Cochran's Q , because these speakers show low rates of verbal behavior under all sources of control. Of course, such cases would still warrant early intensive behavioral intervention.

Statistically significant differences among sources of control over verbal behavior may lead to individualized sequences for prompting and fading. When one operant is statistically stronger than the others, it indicates an individualized type of prompt for conditioning the remaining sources of control. For example, prepotent tact control points to the use of exteroceptive prompts for conditioning mand, echoic, and sequelic sources. Likewise, when one operant is statistically weaker than the others are, it suggests the convergence of multiple sources of control that can then be systematically reduced to the abstracted properties of environmental control. For instance, permutations of tact, echoic, and mand sources can help strengthen, and ultimately isolate, sequelic control.

When used in conjunction with quantitative error analyses, such as those described by Hannula et al. (2020), interventions based on Cochran's Q can offer even further individualization for conditioning the verbal behavior of individuals with ASD. Moreover, Cochran's Q test may be used as a repeated measure to monitor progress over time, and to evaluate the effects of different treatments within single-subject research designs.

In addition to measuring the overselectivity of topography-based verbal behavior, Cochran's Q test may also be a useful measure of restrictive stimulus control over selection-based responding. For instance, the functional analysis could be extended to conditions that include manded-stimulus selection (e.g., presenting multiple stimuli and saying, "Touch ball"), motor imitation (e.g., saying, "Do this," and touching a ball), matching-to-sample (e.g., presenting multiple stimuli and saying, "Match," while displaying a picture of a ball), and selection-by-variable (e.g., presenting multiple stimuli and saying, "Which one do you roll?"). Across each condition, the pointing response is held constant to allow for a comparison of percentages in matched samples.

Likewise, Cochran's Q may be used to measure the extent to which derivational stimulus control is overly restrictive. For example, when analyzing reflexive, symmetric, and transitive relations, Cochran's Q test can be useful for testing the significance of differences between proportions of the three dependent samples.

Extending the Discussion of Overselectivity

Researchers studying stimulus overselectivity have previously attempted to observe this phenomenon within a narrowly defined paradigm bereft of socially valid outcomes.

Here we introduce the use of a well-established statistical analysis, Cochran's Q test, as a means of quantifying stimulus overselectivity within the verbal repertoire. As an inferential statistic, Cochran's Q examines the temporally extended conditioning of various sources of control over verbal behavior, and defines overselectivity as a statistically significant difference between operant classes. The use of Cochran's Q also affords a measure of effect size to quantify the magnitude of restricted stimulus control, and post-hoc analyses to examine specific relationships.

Although none of these statistics tells the complete story of language development, each may be used as a judgmental aid for making data-based instructional decisions (Michael, 1974). With a p -value less than .05, Cochran's Q alerts us to the presence of significantly restricted stimulus control. The chance-corrected effect size, R , quantifies its magnitude according to thresholds that allow us to categorize overselectivity as small, medium, or large. Overall, these numbers tell us whether the operant classes that compose a speaker's verbal repertoire differ to a socially significant degree.

Research on stimulus overselectivity over the past 50 years has become increasingly narrow in scope, leading researchers to make large inferences about the nature of this phenomenon. Lovaas et al. (1971) were aware of the limitations of such a molecular research agenda, arguing, "[a]lthough descriptions of visual attending behavior, which comprise the bulk of research in this area, may provide leads in understanding the psychopathology, such studies are quite inferential" (p. 213). Sidman (1979) stated, "[i]t is not difficult to find instances in which unacknowledged inferences about controlling stimuli have led to impoverished interpretations of complex processes" (p. 134). In light of Sidman's point, ". . . stimulus control is always an inference" (p. 133), we could minimize these particular inferences by assessing overselectivity within a more socially valid context that directly relates to the symptoms associated with ASD.

By examining a temporally extended verbal repertoire via Cochran's Q , we address Gersten's (1980) call to align research on overselectivity with more socially valid outcomes. Future research should also systematically examine Cochran's Q and R with more traditional measures of social validity. For example, changes in Cochran's Q over time could be correlated with parents' descriptions of language ability.

In addition, by incorporating the functional analysis procedures described by Lerman et al. (2005), we answered Gamba et al.'s (2015) call for more rigorous measures of verbal behavior. Future researchers may seek to compare Cochran's Q outcomes against other measures of verbal behavior (e.g., *The Assessment of Basic Language and Learning Skills—Revised* [Partington, 2006]; *Verbal Behavior Milestones Assessment and Placement Program* [Sundberg, 2008]; *Promoting the Emergence of Advanced Knowledge* [Dixon, 2014]; and the *Stimulus Control Ratio Equation* [Mason & Andrews, 2019]) and against other quantitative models of behavior (e.g., the absolute and generalized matching law [Shahan & Podlesnik, 2008]; and log d [Hannula et al., 2020]).

Despite evidence showing that stimulus overselectivity is not unique to individuals with ASD, and not all individuals with ASD exhibit overselectivity, Ploog (2010) proposed that with broadening modifications overselectivity continues to be "a useful conceptualization with implications for a number of the behavioral abnormalities typical for ASD and their treatment, and applicable to the entire heterogeneity of the ASD population" (p. 1345). For half a century, researchers tried unsuccessfully to capture overselectivity as a causal variable for autistic behavior. Operationalizing

stimulus overselectivity with inferential statistics provides a tractable means for its observation and measurement. By distinguishing reliable patterns of restricted stimulus control from aberrations due to random variability, Cochran's Q test allows for more efficacious treatment decisions.

Rincover et al. (1986) referred to overselectivity as a keystone stimulus-control deficit of autism. Although we agree with the sentiment of Rincover's argument, we object to the categorical error. From a behavior-analytic perspective, overselectivity is not a feature, symptom, or cause of ASD. We do not dispute the findings by Reed et al. (2013) and Dube et al. (2016) that typically developing individuals as well as those with other intellectual disabilities, such as Down syndrome, may show restricted stimulus control under certain conditions. When selective responding becomes so pronounced that it occurs to a socially significant degree (i.e., to impair communication and social skills), we begin to call this behavior autistic.

Stokes (1992) explained, "[s]ome social behaviors are best and competently displayed in narrowly defined circumstances, and some are better when shown over a widely diverse set of conditions" (p. 431). Five decades of research have shown that the latter clearly applies to overselectivity, which we argued is more pragmatically conceptualized as differential stimulus control over a population of responses that cannot be attributed to chance alone. Cochran's Q test, along with its corresponding effect-size, can measure the extent to which stimulus overselectivity is both statistically and socially significant.

Acknowledgments The authors thank Jesse Anderson, Duy Le, and Chris Ninness for comments on a prior version of this article.

Declarations

Conflicts of interest We have no explicit conflicts of interest to disclose.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43(3), 666–678. <https://doi.org/10.3758/s13428-011-0089-5>
- Baron-Cohen, S. (2005). Autism—"autos": Literally, a total focus on the self? In T. E. Feinberg & J. P. Keenan (Eds.), *The lost self: Pathologies of the brain and identity* (pp. 166–180). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195173413.003.0011>
- Berry, K. J., Johnston, J. E., & Mielke Jr., P. W. (2007). An alternative measure of effect size for Cochran's Q test for related proportions. *Perceptual & Motor Skills*, 104, 1236–1242. <https://doi.org/10.2466/pms.104.4.1236-1242>
- Brown, S. M., & Bebko, J. M. (2012). Generalization, overselectivity, and discrimination in the autism phenotype: A review. *Research in Autism Spectrum Disorders*, 6(2), 733–740. <https://doi.org/10.1016/j.rasd.2011.10.012>
- Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37(3–4), 256–266. <https://doi.org/10.2307/2332378>
- Cohen, B. H. (2013). *Explaining psychological statistics* (4th ed.). John Wiley and Sons.

- Davis, D. G. (2017). *Novice behavior in a makerspace: Pragmatic pathways to shaping CS identity* (Doctoral dissertation, University of Texas at San Antonio).
- Davison, M. (1999). Statistical inference in behavior analysis: Having my cake and eating it? *The Behavior Analyst*, 22(2), 99–103. <https://doi.org/10.1007/BF03391986>
- Dixon, M. R. (2014). *The PEAK relational training system module 1: Direct training*. Shawnee Scientific Press.
- Dowdy, A., Peltier, C., Tincani, M., Schneider, W. J., Hantula, D. A., & Travers, J. C. (2021). Meta-analyses and effect sizes in applied behavior analysis: A review and discussion. *Journal of Applied Behavior Analysis*. <https://doi.org/10.1002/jaba.862>
- Dube, W. V., Farber, R. S., Mueller, M. R., Grant, E., Lorin, L., & Deutsch, C. K. (2016). Stimulus overselectivity in autism, Down syndrome, and typical development. *American Journal on Intellectual & Developmental Disabilities*, 121(3), 219–235. <https://doi.org/10.1352/1944-7558-121.3.219>
- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241–252. <https://doi.org/10.1080/00401706.1964.10490181>
- Fairbanks, D. J., & Rytting, B. (2001). Mendelian controversies: A botanical and historical review. *American Journal of Botany*, 88(5), 737–752. <https://doi.org/10.2307/2657027>
- Ferster, C. B., Culbertson, S., & Boren, M. C. (1975). *Behavior principles* (2nd ed.). Prentice Hall.
- Gamba, J., Goyos, C., & Petursdottir, A. I. (2015). The functional independence of mands and tacts: Has it been demonstrated empirically? *Analysis of Verbal Behavior*, 31(1), 10–38. <https://doi.org/10.1007/s40616-014-0026-7>
- Gersten, R. M. (1980). In search of the cognitive deficit in autism: Beyond the stimulus overselectivity model. *Journal of Special Education*, 14(1), 47–65. <https://doi.org/10.1177/002246698001400106>
- Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51(1), 79–88. <https://doi.org/10.1037/h0046219>
- Hall, G., & Sundberg, M. L. (1987). Teaching mands by manipulating conditioned establishing operations. *Analysis of Verbal Behavior*, 5(1), 41–53. <https://doi.org/10.1007/bf03392819>
- Hannula, C., Jimenez-Gomez, C., Wu, W., Brewer, A. T., Kodak, T., Gilroy, S. P., Hutsel, B. A., Alsop, B., & Podlesnik, C. A. (2020). Quantifying errors of bias and discriminability in conditional-discrimination performance in children diagnosed with autism spectrum disorder. *Learning & Motivation*, 71, Article 101659. <https://doi.org/10.1016/j.lmot.2020.101659>
- Hoaglin, D. C. (2016). Misunderstandings about Q and “Cochran's Q test” in meta-analysis. *Statistics in Medicine*, 35(4), 485–495. <https://doi.org/10.1002/sim.6632>
- Honig, W. K., & Urcioli, P. J. (1981). The legacy of Guttman and Kalish (1956): Twenty-five years of stimulus generalization research. *Journal of the Experimental Analysis of Behavior*, 36, 405–445. <https://doi.org/10.1901/jeab.1981.36-405>
- Ibrahim, K., & Sukhodolsky, D. G. (2018). RDoC and autism. In F. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders*. Springer. https://doi.org/10.1007/978-1-4614-6435-8_102261-1
- Kelley, M. E., Shillingsburg, M. A., Castro, M. J., Addison, L. R., LaRue Jr., R. H., & Martins, M. P. (2007). Assessment of the functions of vocal behavior in children with developmental disabilities: A replication. *Journal of Applied Behavior Analysis*, 40(3), 571–576. <https://doi.org/10.1901/jaba.2007.40-571>
- Kelly, M. P., & Reed, P. (2020). Examination of stimulus over-selectivity in children with autism spectrum disorder and its relationship to stereotyped behaviors and cognitive flexibility. *Focus on Autism & Other Developmental Disabilities*, 36(1), 47–56. <https://doi.org/10.1177/1088357620943504>
- Kyonka, E. G. (2019). Tutorial: Small-N power analysis. *Perspectives on Behavior Science*, 42(1), 133–152. <https://doi.org/10.1007/s40614-018-0167-4>
- LaFrance, D., Wilder, D. A., Normand, M. P., & Squires, J. L. (2009). Extending the assessment of functions of vocalizations in children with limited verbal repertoires. *Analysis of Verbal Behavior*, 25, 19–32. <https://doi.org/10.1007/BF03393067>
- Lerman, D. C., Parten, M., Addison, L. R., Vomdran, C. M., Volkert, V. M., & Kodak, T. (2005). A methodology for assessing the functions of emerging speech in children with developmental disabilities. *Journal of Applied Behavior Analysis*, 38(3), 303–316. <https://doi.org/10.1901/jaba.2005.106-04>
- Lovaas, O. I., & Schreibman, L. (1971). Stimulus overselectivity of autistic children in a two stimulus situation. *Behaviour Research & Therapy*, 9, 305–310. [https://doi.org/10.1016/0005-7967\(71\)90042-8](https://doi.org/10.1016/0005-7967(71)90042-8)
- Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin*, 86, 1236–1254. <https://doi.org/10.1037/0033-2909.86.6.1236>
- Mason, L. L., & Andrews, A. (2019). The verbal behavior stimulus control ratio equation: A quantification of language. *Perspectives on Behavior Science*, 42(2), 323–343. <https://doi.org/10.1007/s40614-018-0141-1>
- McNemar, Q. (1962). *Psychological statistics* (3rd ed.). John Wiley & Sons.

- Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, 7(4), 647–653. <https://doi.org/10.1901/jaba.1974.7-647>
- Moxley, R. (1987). Three conceptual units for behavior. *The Behavior Analyst*, 10(1), 17–26. <https://doi.org/10.1007/bf03392403>
- Normand, M. P., Severson, E. S., & Beavers, G. A. (2008). A functional analysis of nonvocal verbal behavior of a young child with autism. *Analysis of Verbal Behavior*, 24, 63–67. <https://doi.org/10.1007/BF03393057>
- Partington, J. W. (2006). *The assessment of basic language and learning skills—revised (ABLLS–R)*. Behavior Analysts.
- Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism & Developmental Disorders*, 40, 1332–1349. <https://doi.org/10.1007/s10803-010-0990-2>
- Reed, S. R., Stahmer, A. C., Suhrheinrich, J., & Schreibman, L. (2013). Stimulus overselectivity in typical development: Implications for teaching children with autism. *Journal of Autism & Developmental Disorders*, 43, 1249–1257. <https://doi.org/10.1007/s10803-012-1658-x>
- Rieth, S. R., Stahmer, A. C., Suhrheinrich, J., & Schreibman, L. (2015). Examination of the prevalence of stimulus overselectivity in children with ASD. *Journal of Applied Behavior Analysis*, 48(1), 71–84. <https://doi.org/10.1002/jaba.165>
- Rincover, A., Feldman, M., & Eason, L. (1986). “Tunnel vision”: A possible keystone stimulus control deficit in autistic children. *Analysis & Intervention in Developmental Disabilities*, 6(4), 283–304. [https://doi.org/10.1016/s0270-4684\(86\)80010-6](https://doi.org/10.1016/s0270-4684(86)80010-6)
- Shahan, T. A., & Podlesnik, C. A. (2008). Quantitative analyses of observing and attending. *Behavioural Processes*, 78(2), 145–157. <https://doi.org/10.1016/j.beproc.2008.01.012>
- Sidman, M. (1979). Remarks. *Behaviorism*, 7, 123–126. <https://doi.org/10.1080/15021149.2011.11434377>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
- Skinner, B. F. (1973). Answers for my critics. In H. Wheeler (Ed.), *Beyond the punitive society* (pp. 256–266). W. H. Freeman.
- Stokes, T. (1992). Discrimination and generalization. *Journal of Applied Behavior Analysis*, 25(2), 429–432. <https://doi.org/10.1901/jaba.1992.25-429>
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10(2), 349–367. <https://doi.org/10.1901/jaba.1977.10-349>
- Sundberg, M. L. (2007). *The Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP): Field-test data from typical children and children with autism*. Paper presented at the meeting of the Association for Behavior Analysis International, San Diego, CA.
- Sundberg, M. L. (2008). *Verbal Behavior Milestones Assessment and Placement Program: The VB-MAPP*. AVB Press.
- Tate, M. W., & Brown, S. M. (1970). Note on the Cochran Q test. *Journal of the American Statistical Association*, 65(329), 155–160. <https://doi.org/10.1080/01621459.1970.10481069>
- Wing, L. (1988). The continuum of autistic disorders. In E. Schopler & G. M. Mesihov (Eds.), *Diagnosis and assessment in autism* (pp. 91–110). Plenum.
- Young, M. E. (2018). A place for statistics in behavior analysis. *Behavior Analysis: Research & Practice*, 18(2), 193–202. <https://doi.org/10.1037/bar0000099>

Affiliations

Lee Mason^{1,2} · **Maria Otero**^{1,3} · **Alonzo Andrews**^{4,5}

¹ Child Study Center, Cook Children's Health Care System, 1300 West Lancaster Avenue, Fort Worth, TX 76102, USA

² Texas Christian University, Fort Worth, TX, USA

³ University of North Texas, Denton, TX, USA

⁴ University of Texas at San Antonio, San Antonio, TX, USA

⁵ San Antonio Independent School District, San Antonio, TX, USA