MEDWIN PUBLISHERS
Committed to create value for Researchers

Psychology & Psychological Research International Journal
ISSN: 2576-0319

# Stimulus Fading Throughout the Years and its Impact on Children's Discrimination Learning: A Meta-Analysis

**Maag JW***

Department of Special Education and Communication Disorders, University of Nebraska-Lincoln, USA

***Corresponding author:** John W Maag, Department of Special Education and Communication Disorders, 202 Barkley Memorial Center, University of Nebraska-Lincoln, Lincoln, NE 68583, USA, Tel: 4024725477; Email: jmaag1@unl.edu

**Review Article**

Volume 4 Issue 4
**Received Date**: June 10, 2019
**Published Date:** July 16, 2019
**DOI:** 10.23880/pprij-16000211

## Abstract

Many children with intellectual or developmental disabilities and those with autism experience difficulty learning and often need more salient cues to prompt correct responding. However, in order for generalization to occur, behaviors need to become prompted by less obvious more natural environmental stimuli. Stimulus fading is a method to accomplish this goal. It involves gradually reducing the level and degree of stimuli required to cue acquisition of discrimination learning. It has been used for over 40 years, yet no systematic review has been conducted to determine its effectiveness. The purpose of the current study was to conduct a meta-analysis of the literature on stimulus fading with children and adolescents. Effect sizes were calculated using standard mean difference (SMD), improvement rate difference (IRD), and Tau-*U*. The obtained effect values indicated that stimulus fading is a moderately effective method, with some effect sizes in the highly effective range, for helping children learn and perform skills. Considerations and variables regarding the effectiveness of stimulus fading are discussed.

**Keywords:** Discrimination Learning; Applied Behavior Analysis; Fading; Single Case Research Design; Autism Spectrum Disorders

## Introduction

Stimulus fading is one of several techniques to promote errorless learning most commonly used with individuals with autism, intellectual disabilities, or pervasive developmental disabilities [1]. It is based on Skinner's [2] operant model of behavior and operationalized within the context of applied behavior analysis and whose dimensions were described by Baer, Wolf, and Risley [3]. Stimulus fading specifically refers to the gradual change of the stimulus controlling a behavior, such that the behavior eventually occurs as the result of a partially changed or completely new antecedent stimulus. It is used to transfer stimulus control from a prompt to the natural stimulus. Any number and types of antecedents lend themselves to stimulus fading such as color, sound, or size. For example, a child could be taught to discriminate between the colors blue and red by first presenting a blue beach ball and a red tennis ball. The larger size of the blue ball provides a more salient stimulus to evoke the correct response. The blue ball would gradually be made successively smaller (e.g., blue soccer ball then blue softball) until both a blue and red tennis ball would be presented. Stimulus fading was a departure from the previous approach of trial and error [4] in which the correct choice (blue tennis ball = S+) and error choice (red tennis ball = S-) were presented simultaneously.

There are two variables to consider when using stimulus fading [5]. The first variable is how different are a stimulus fading step from the previous step. Typically, stimulus hierarchies begin with small steps leading to the terminal behavior [6,7] rather than generating them for each step [8]. A second variable is determining how quickly to move through the steps in a stimulus fading hierarchy. Some researchers gradually move through each step in the hierarchy [9,10] while others occasionally probe the terminal goal during the hierarchy [11,12]. Sometimes pre-instructional probes are used to determine the initial stimulus fading step [13], but these probes typically are not used throughout the stimulus fading process [6,7]. Regardless, it is difficult to determine if a person would need to be exposed to all or only some of the steps in a stimulus fading hierarchy. Schiff, et al. [7] found systematic probes that were presented three steps ahead of the current step eliminated about 25% of the stimulus fading steps necessary for successful discrimination learning to occur.

The first experimental study of stimulus fading was conducted by Terrace [14] using pigeons who were trained in a successive discrimination task in which key pecks to a red light (S+) were reinforced and those to a green light (S-) were not. Stimulus control (i.e., red light) was established without the green light instead of both lights appearing together at their final strength which would create a situation in which errors could have been made by selecting the green light. Essentially, key pecks in the presence of the red light were reinforced. The green light's duration and intensity were gradually faded beginning with a very short and low presentation until both lights were identical, but only after consistent responding was established and key pecks were under discriminative control of the red light. The pigeons did not peck the green key when it was presented in its final form, thus indicating discriminative control from the red light was established in the initial absence of the green light. The incorrect choice (S) was gradually added (i.e., green light) to generate opportunities for errors.

Terrance [15] conducted a follow-up study, also with pigeons, showing how differential reinforcement in the presence of two or more discriminative stimuli created a contrast with shorter latencies of responding and that an essential provision for contrast is the occurrence of responses to S- while acquiring a discrimination (S+). He conducted two experiments. In experiment 1, the discriminative stimuli (i.e., red-green lights and vertical-horizontal lines) trials were presented briefly, discretely, and automatically. In experiment 2 the only procedural difference from experiment 1 was the number of training sessions. Results from both experiments were that pigeons could learn an easy discrimination of color and a more difficult discrimination of the orientation of a line without making any errors.

Arguably, the first study to examine stimulus fading with children was Schreibman [16]. A prompt-fading approach was used to facilitate discrimination with six children with autism (four male and two female). Two of the tasks involved visual stimuli (forms on cards) and two involved auditory stimuli (two-syllable nonsense words). There were three conditions: (1) training without a prompt, (2) extra-stimulus prompt fading, and (3) within-stimulus prompt fading. There were three main results. First, participants were typically unsuccessful learning the discriminations without a prompt. Second, participants were unable to learn a discrimination when given the extra-stimulus prompt but were successful when the within-stimulus prompt was presented. Third, results were the same across participants regardless of the prompt being auditory or visual.

There have been many research studies examining the effectiveness of various forms of stimulus fading since 1975. The most recent study to date was conducted by Valentino, LeBlanc, and Raetz [17]. The participant was a 10-year old boy with autism and his problem was consuming food at extremely high rates. The goal was to slow down his rate by using a vibrating pager to cue him that he was eating too rapidly. Two types of fading of the pager were used. First, fading the frequency of the vibrations was implemented and found to be ineffective. However, the second technique, fading the intensity of the vibrations, was successful.

It is quite astonishing the range of situations and behaviors that have been addressed with stimulus fading. Many of the studies focus on various aspects of medical procedures for children with autism. For example, two studies taught children with autism to comply with needle injections [9,12]. Cuvo and his colleagues treated children with autism be compliant with oral and physical examinations [6,10]. Other medical related problems such as pill-swallowing difficulties and compliance with liquid medication administration have been treated with stimulus fading [7,13,18]. Applications of stimulus fading have also been used to engender compliance with toothbrushing, food refusal, elective mutism, and self-injurious behaviors [5,11,19,20].

Given the number of decades stimulus fading has been used and the wide array of applications, it is somewhat surprising that in between Schreibman's [16] study and the one by Valentino, et al. [17] there does not appear that any systematic reviews—narrative or meta-analytic—have been conducted on the stimulus fading literature. There have been some non-systematic literature reviews that have included stimulus fading. Mueller, et al. [1] conducted a literature review of various errorless learning procedures in which one was stimulus fading. Similarly, Vriend, Corkum, Moon, and Smith [21] reviewed behavioral interventions for sleep problems in children with autism that has a section on stimulus fading. Three reviews focused on the treatment of elective mutism and identified stimulus fading as one of several interventions, although they are quite dated [22-24]. Therefore, the purpose of the current study was to conduct a systematic meta-analytic review of the stimulus fading literature for children. A second purpose was to evaluate the quality of studies reviewed by applying the eight quality indicators encompassing 22 items developed by the Council for Exceptional Children [25] for single case research design (SCRD) studies.

## Method

A systematic search was performed to identify the extant research regarding the use of stimulus fading for children and adolescents. The search methods were consistent with the 12-item PRISMA statement for reporting meta-analyses [26]. The purpose was to ensure clarity and transparency of conducting systematic reviews.

Academic Search Premier was the search source with the following selected databases: ERIC, MedLINE, PsycARTICLES, and PsycINFO. The following Boolean terms/phrases were used: ("stimulus fading") AND ("child") OR ("adolescents") OR ("youth") OR ("teenagers") OR ("students") OR ("students with disabilities"). In addition, ancestral searches were conducted of four journals that publish exclusively or primarily SCRD studies: *Journal of Applied Behavior Analysis*, *Journal of Behavioral Education*, *Research in Autism Spectrum Disorders*, *Education and Treatment of Children*, *Journal of Autism and Developmental Disorders*, and *Behavioral Interventions*.

### Eligibility Criteria and Study Selection

Studies included were those only using SCRDs, had to be in English, and published in peer-reviewed journals between January 1, 1963 and December 30, 2018. The start year of 1963 was selected because it was when the first experimental study of stimulus fading was conducted [14]. Participants had to be between the ages of 4 and 18. It did not matter what setting the studies were conducted.

Studies were identified and retained at different stages based on PRISMA guidelines, and the results are displayed in Figure 1. There were 127 total records identified that were articles appearing in peer-reviewed journals. Of those, 26 were read in their entirety (i.e., method sections for inclusion criteria). Two graduate students were trained by the researcher how to read each of the 26 studies method sections. One graduate student read all 26 studies (i.e., method sections) while the other read 10 randomly selected studies and their interrater agreement was 100%. After engaging in the flow of information process, there were 13 articles retained for the current review.

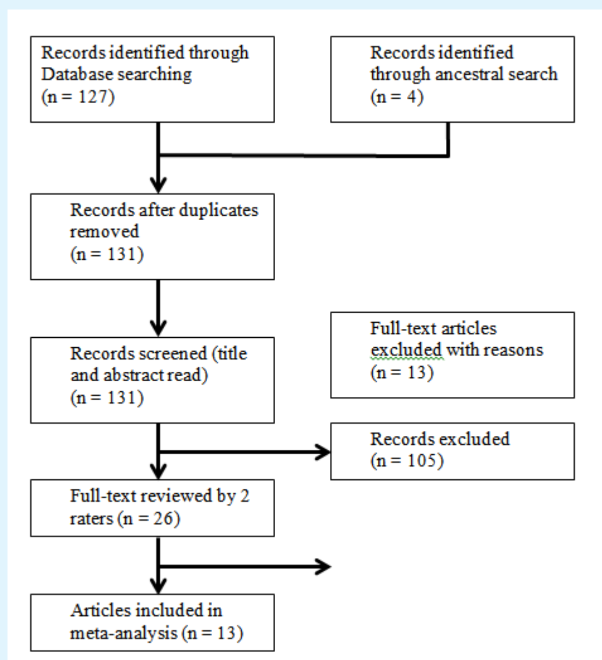


**Figure 1:** Search Results using PRISMA Guidelines.

## Coding Procedures

**Descriptive characteristics**: The 13 articles retained from the search were coded along five variables: (a) participant age, gender, and, diagnosis/educational label; (b) setting; (c) type of design; (d) dependent variables; and (e) type of fading procedure. Two graduate assistants were trained by the experimenter to code the five variables. Four studies were randomly selected and the experimenter demonstrated the coding process on two through instructions and modeling. The two graduate assistants coded the remaining two studies with the experimenter providing performance feedback. The two graduate students then each coded the remaining studies independent of each other. Interrater reliability (IRR) was calculated for seven randomly selected studies (50%). This percentage was congruent with other published SCRD meta-analyses [27-29].

**Methodological quality**: The same two graduate assistants appraised the quality of each article based on the Council for Exceptional Children's (CEC) *Standards for Evidence-Based Practices* (2014) that consisted of 22 component items across eight quality indicators (QIs) for SCRDs. The same training format used for coding descriptive characteristics was used for coding QIs. A binary score of one (met) or zero (not met) were used in the coding scheme (i.e., absolute coding) rather than using weighted coding [30]. The reason for using the more stringent binary approach was to require raters to make more unequivocal decisions rather then being uncertain and overusing the .5 (partially met) criteria.

A coding sheet with the 22 components across eight QIs for SCRD studies was created in Excel©. The sheet consisted of three columns. The first column contained the QI, the second column had the description, and the third column consisted of clarification developed by Common et al. [30].

## Statistical Analysis

**Data extraction:** Data were extracted from graphs in each study using Enguage Digitizer [31] -an open source digitizing software package that converts graphic image files (e.g., .jpg, .bmp) into numerical data. Enguage is a free software package that is comparable to Biosoft's Ungraph 5.0 that was recommended in the manual developed by Nagler, Rindskopf, and Shadish [32] for conducting SCRD meta-analyses and used in previous meta-analyses [28]. In addition, Losinski and his colleagues also converted all scores into percentages setting the upper level and lower level of the *y* axis on all students to 100 and 0, respectively, before extraction. Their rational was to address (a) the inherent subjectivity in which target variables were operationally defined, for example, "aggression" versus "hitting and pushing" or different behaviors that make up "off-task" and (b) the capricious nature of measurement on the y axis (e.g., wide ranges between numbers, measurement differences such as frequency versus duration, or percentages).

**Effect size calculations:** Horner, Swaminathan, Sugai, and Smolkowski [33] noted that currently there is no consensus for the method for quantifying outcomes for use in SCRD meta-analysis, although some effect size calculations may be more robust or appropriate than others depending on data characteristics. Therefore, three types of effect sizes were calculated in the present meta-analysis. Standard mean difference (SMD) was calculated because it is the SCRD analog or variation of Cohen's *d* statistic where the mean of the baseline phase is subtracted from the mean of the intervention phase and divided by the pooled standard deviation [34]. The similarity to Cohen's *d* makes SMD an important statistic for comparison to non-single-case methods. However, SMD is considered by some unreliable because of small number of observations and floor effects limiting variability and results in overestimates of the parametric treatment effects [33,35]. Losinski, et al. [28] addressed this problem by establishing a ceiling value of *d* at the 3rd quartile of the total distribution in order to account for statistical outliers disproportionally affecting the outcomes of the studies when aggregated. This practice (i.e., 3rd quartile ceiling) was also used in the present study resulting in a ceiling of $d = 2.44$. Improvement rate difference (IRD) was also computed because it provides an effect size similar to the *risk difference* used in medical treatment research which has a proven track record in hundreds of studies [36]. Finally Tau-*U* values were computed because it controls for monotonic trend (i.e., increasing data during baseline). The IRD and Tau-*U* were calculated using the www.singlecaseresearch.org/calculators. For studies in which improvement was in the decreasing direction, the correction feature for Tau-*U* was not used (i.e., only Tau). Effect sizes for SMD were computed using a calculator.

**Publication Bias:** Publication bias, or the "file drawer" effect refers to presence of potential bias existing because of a greater likelihood that published research shows positive findings [37]. In a meta-analysis of group design studies, the Meta-Win's Fail-Safe function [38] can be used to estimate the number of studies with null results sufficient to reduce observed effect size to a minimal level (i.e., < .20). However, there is no comparable formula in SCRD meta-analyses. Therefore, to reduce the likelihood of the "file drawer" effect, the number of cases with no effect (i.e., 0) were added to the group of study effect sizes to reduce the overall effect to insignificant or suspect levels ($d$ <.20; IRD < .37; Tau <.20). Effect sizes calculated for SMD were not used in the process because of their tendency to overestimate results. This process results with the number of participants in potentially "filed" studies (i.e., not submitted for publication for whatever reasons) needed to reduce effect sizes of included studies to insignificant levels (i.e., no observed effect).

**Interrater Reliability:** Interrater reliability (IRR) data were conducted on six randomly selected articles out of the 13 included studies for a total of 46% of studies on the five coded study characteristics and eight (22 items) QIs. Interrater reliability was calculated both for study characteristics and QI components by dividing the total number of agreements by the total number of agreements plus disagreements for each item and averaged for all items. Two graduate research assistants coded the articles for all variables and IRR for study characteristics was 84% (range: 58% -100%; SD=15.819) and 82.8% (range 64% - 91%; SD=11.166) for QIs.

## Results

Results are presented in three sections. The first section addresses descriptive features obtained from the studies including characteristics of participants and settings, design features, dependent variables, and fading technique. The second section presents the extent to which studies met each of the 22 component items pertaining to SCRD across CEC's eight QIs. The final section contains the statistical analysis of effect sizes results of publication bias.

### Descriptive Features of Included Studies

**Characteristics of participants and settings:** A total of 34 participants (27 male and 7 female) were included in the 13 studies contained in this analysis. Participants' ages ranged from four years old [5,6,10] to 18 years old [12] with an average of 8.74 for males (range = 4 – 18, SD = 4.043) and 8.71 for females (range = 5 – 15, SD = 3.817). There were a total of six participants who were excluded from the analysis because of being younger than four years of age [6,10,20]. The majority of participants (n=29) were diagnosed with autism or autism spectrum disorder (ASD). Four participants were diagnosed with a developmental disability [10,39] and seven had comorbid conditions [11,12,20,40].

Included studies took place in a variety of settings—most related to the dependent variable of interest. For example, the Bishop, et al. [5] study took place in the bathroom of the participant's home because the goal was to increase the percentage of correct tooth brushing. A dental clinic and medical office on an university campus served as the settings for Cuvo, Godard, et al. [6] and Cuvo, Reagan, et al. [10] because of the dependent variables of being compliant with a dental hygienist and during a physical examination, respectively. For similar reasons, the setting for the Valentino, et al. [17] study was the kitchen of the participant's school because of trying to increase the total number of bites of food. Four studies took place in participants' classroom and a common theme was the dependent variable being some type of increase in verbal interaction [9,40-42]. Table 1 contains the settings, dependent variables, and fading techniques.

| Study | Dependent Variable(s) | Setting | Fading Technique |
|---|---|---|---|
| 1. Birkan, et al. [9] | Correct responses on each task without stereotypy or disruptive behavior (e.g., screaming, grabbing materials, biting) to cooperate with injections. | Participants classroom when other children were absent. | Fading in an aversive stimulus (i.e., injection) Ice cube was used as a stimulus, then fading in a paper clip, then the actual injection |
| 2. Bishop, et al. [5] | Percent of correct toothbrushing | Child's home in a bathroom where toothbrushing typically occurred | 30-step stimulus fading hierarchy by gradually increasing the proximity of the toothbrush to the child's mouth. Final stimulus: tooth brushing for 60 seconds |
| 3. Brown, et al. [41] | Frequency of verbal interactions per minute during simulated shopping | School classroom with a carpeted | Script fading (last word to first word) to cue verbal interactions. Final stimulus: unscripted |

| | | | |
|---|---|---|---|
| | trips & during visits to community stores. | floor reconfigured to create three mock stores. | statements |
| 4. Cuvo, Godard, et al. [6] | Compliance to dental hygienist (DH): (a) participant tolerated DH performing procedures without emitting escape/avoidance behaviors, or (b) emitted a response within 10 s when prompted without escape/avoidance behavior. Aggression and property destruction; whining or crying. Disruption (e.g., stereotypic behavior, lifting and dropping limbs on dental chair). | Dental clinic on a university campus. | Stimulus to cue behavior: mirror that was faded in by bringing in closer proximity to the child<br>Final Stimulus: participants opened their mouth when instructed, and tolerated the mirror in their mouth for 10 s |
| 5. Cuvo, Reagan, et al. [10] | Frequency of 10 compliance steps without protesting, aggression, property destruction, or escape/avoidance behaviors during a physical examination | Medical office in the physician assistant program on university campus | Proximity to the aversive stimuli (e.g., stethoscope) was faded in while the participants had access to preferred stimuli. |
| 6. Ducharme & Worling [39] | Percentage of compliance to low-probability "do" and "don't" requests. | Family home | Systematic reduction in the number of high probability requests (three to two to one) and increased latency between the high- and low-probability |
| 7. Freeman & Piazza [11] | Number of grams of food consumed and compliance (consumption of a bite following a verbal or partial physical prompt) for each meal | Hospital | Amount of presented food was increased in 5% increments of the age-appropriate portions provided by the hospital. Final Stimulus: Participant was consuming 50% of an age-appropriate meal |
| 8. Pace, et al. [20] | Self-injurious behavior: self-hitting, forceful contact of the hand or fist against the face, head, or legs, hand biting, and scratching | Therapy room equipped with furniture, materials, and one-way observation window | Instructional trials faded in the absence of self-injurious behavior |
| 9. Sarokoff, et al. [40] | Number of scripted and unscripted verbal statements identical to the written script; unprompted statements not present in the script | Classroom, treatment room, and activities room | Five-step script-fading procedure to fade out different words within the sentence until only the package (food or game) was present.<br>Step 1, 25% of the words faded from back to front; Step 2, half of each sentence faded; Step 3, the package and first letter of each line remained; Step 4, the paper was presented with the package; and in Step 5, only the package remained. |
| 10. Schreibman | Percent of correct verbal responses; pointing to the correct stimulus card | Children's treatment Center at a state | Two different cards, a pointing prompt was given and faded by the experimenter moving |

| | | | |
|---|---|---|---|
| [16] | | hospital | her finger along the side of the card until the prompt was completely faded and correct responses were obtained with no finger prompt.<br>Within-stimulus fading: visually discriminate difference between two cards. The second card blank and then made increasingly dark until it was same color intensity as the correct card.<br>Within stimulus prompt fading emphasizing auditory differences between S+ and S- |
| 11. Shabani & Fisher [12] | Percentage of successful trials drawing blood without pulling away; moving arm no more than 3 cm from the outline of where his arm could be in any direction during a 10 s trial | Treatment room at outpatient clinic | The lacent for drawing blood was started far away from participant's arm and the distance of from the correct placement on his arm was faded until drawing blood was successful. |
| 12. Taylor & Levin [42] | Frequency of verbal initiations during play and follow-up sessions without verbal models related to the activity (e.g., "look at this truck"), were directed toward another person (e.g., used the person's name, said "look," or oriented toward the person), and were complete sentences. | Classroom after school hours | Teacher verbal prompts cued by vibrating device decreased until a verbal initiation was made each time the device vibrated. |
| 13. Valentino, et al. [17] | Number of seconds of eating time to consume strawberries; total number of bites | Kitchen of the school participant attended. | Pager prompt to cue participant to take a bite. Intensity and frequency of the prompts reduced until pager was not activated. |

**Table 1:** Dependent variable, setting, and fading technique.

**Design features:** A multiple baseline design across participants was used in three studies [5,40,41] and two more using the similar multiple probe design [6,10]. Three studies used a multi-element design [11,20,42] while two others used changing criterion and changing conditions designs [9,16]. These three designs are common among stimulus fading studies because of the different phases, depending on the number of individual fading steps, being evaluated. Two studies used variations of the reversal design [12,17] while Ducharme and Worling [39] embedded a reversal design within a multiple baseline design.

**Dependent variables:** The type of dependent variables targeted varied greatly from study to study primarily because of the versatility of behaviors to which stimulus fading can be applied. Nevertheless, there were three common themes for which stimulus fading was applied: (a) medical or dental procedures [9,6,10,12], (b) food intake and personal hygiene [5,11,17] and (c) social interactions [40-42].

**Types of fading procedures:** The type of fading procedure used obviously differed depending on the dependent variable of interest. The most common approach was to develop stimulus fading hierarchies (n = 5) with the number of steps ranging from seven [41] to 30 [5]. More creative but unscripted approaches were used such as Birkan, et al. [9] increasing cooperation with injections with the initial stimulus touching participants with an ice cube, then a straightened paper clip, and finally the actual injection. Two studies used vibration devices and then faded either or both the duration between vibrations and intensity of the vibrations [17,42].

**Methodological Quality Indicators:** CECs *Standards for Evidence-Based Practices* (2014) that consisted of 22 component items across eight quality indicators (QIs) for SCRDs were used to determine methodological quality of reviewed studies. One study met all 22 items [6] and one study met 21 [10]. The lowest score (15) was obtained for five studies [9,11,12,17,41]. Overall, quality of the 13 studies was moderate (mean=16.92, SD=2.465, range 15–22). The lowest score (n=4) was for all three items (5.1, 5.2, 5.3) under implementation fidelity and 6.1 *Independent variable (IV) systematically manipulated*. There were eight items met by all studies: 1.1 *Context/setting description*, 3.2 *Training description*, 6.4 *Design demonstrates experimental effect at three different times*, 6.5 *Baseline contains at least three data points*, 6.6 *Design controls for threats to internal validity*, 7.3 *Reports effects of the intervention on all measures*, 7.5 *Adequate interobserver agreement*, and 8.1 *Study provides single-case graph data across all study phases*. Table 2 displays the results for each component across QIs for all 13 studies. The grey shaded cells are those items that received a 0 (n=67 [24%]). Taken together, only four out of 13 studies (31%) met 80% or more of the QI items (m=77%, SD = 11.119, range=68% - 100%).

| | **Quality Indicators (22 Items)** | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Study** | **1.1** | **2.1** | **2.2** | **3.1** | **3.2** | **4.1** | **4.2** | **5.1** | **5.2** | **5.3** | **6.1** | **6.2** | **6.3** | **6.4** | **6.5** | **6.6** | **7.1** | **7.2** | **7.3** | **7.4** | **7.5** | **8.1** | **Total** |
| 1. Birkan, et al. [9] | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | **15** |
| 2. Bishop, et al. [5] | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | **16** |
| 3. Brown, et al. [41] | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **15** |
| 4. Cuvo, Godard, et al. [6] | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **22** |
| 5. Cuvo, Reagan, et al. [10] | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **21** |
| 6. Ducharme & Worling [39] | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **19** |
| 7. Freeman & Piazza [11] | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | **15** |
| 8. Pace, et al. [20] | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **19** |
| 9. Sarakoff , et al. [40] | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **16** |
| 10.Schreibman [16] | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **16** |
| 11. Shabani & Fisher [12] | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | **15** |
| 12. Taylor & Levin [42] | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **16** |
| 13.Valentino, et al. [17] | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | **15** |
| **Total** | **13** | **12** | **8** | **8** | **5** | **13** | **12** | **4** | **4** | **4** | **4** | **11** | **10** | **12** | **13** | **13** | **10** | **12** | **13** | **12** | **13** | **13** | |

*Note***.** 1.1 Context/setting description; 2.1 Participant description, 2.2 Participant disability/at-risk status; 3.1 Role and description; 3.2 Training description; 4.1 Intervention procedures; 4.2 Materials description; 5.1 Implementation fidelity assessed/reported; 5.2 Fidelity dosage or exposure assessed/reported; 5.3; Fidelity assessed across relevant elements/throughout study; 6.1 Independent variable (IV) systematically manipulated; 6.2 Baseline description; 6.3 No or limited access to IV during baseline; 6.4 Design provides at least demonstrations of experimental effect at three different times; 6.5 Baseline phase contains at least three data points; 6.6 Design controls for common threats to internal validity; 7.1 Socially important goals; 7.2 Description of dependent variable measures; 7.3 Reports effects of the intervention on all measures; 7.4 Minimum of three data points per phase; 7.5 Adequate interobserver agreement; 8.1 Study provide single-case graph clearly representing outcome data across all study phases

**Table 2:** Quality Indicators Met by Study.

### Statistical Analysis

**Effects of studies**: Effect sizes were calculated for each AB contrast. Several studies used a multiple baseline design across three behaviors so those participants would have three AB contrasts. Effect sizes were then averaged for each study and appear in Table 3. Overall omnibus effect sizes for each type were as follows: SMD (mean = 2.01, SD = 0.403, range = 1.36 – 2.44); IRD (mean = .813, SD = 0.189, range = 0.43 – 1); and Tau-*U*, mean = .547 (SD = 0.677, range = -1 – 1).

| Study | Effect Sizes | | | | | |
|---|---|---|---|---|---|---|
| | SMD | | IRD | | Tau-*U* | |
| | Mean | SD | Mean | SD | Mean | SD |
| 1. Birkan et al. (2011) [a] | N/A | N/A | N/A | N/A | N/A | N/A |
| 2. Bishop et al. (2013) | 1.668 | 0.195 | 0.728 | 0.100 | 0.718 | 0.102 |
| 3. Brown et al. (2008) [a] | 2.014 | N/A | 0.904 | 0.121 | 0.921 | 0.122 |
| 4. Cuvo, Godard, et al. (2010) [b] | 1.755 | 0.435 | 0.793 | 0.135 | 0.886 | 0.122 |
| 5. Cuvo, Reagan, et al. (2010) [b] | 1.376 | 0.451 | 0.781 | 0.165 | 0.811 | 0.138 |
| 6. Ducharme & Worling (1994) | 1.928 | 0.896 | 0.683 | 0.340 | 0.729 | 0.341 |
| 7. Freeman & Piazza (1998) [a c] | N/A | N/A | N/A | N/A | N/A | N/A |
| 8. Pace et al. (1993) [c d] | 2.436 | 0 | 1 | 0 | -1 | 0 |
| 9. Sarokoff et al. (2001) [b] | 2.436 | 0 | 8.320 | 0.890 | 1 | 0 |
| 10. Schreibman (1975) [a] | N/A | N/A | N/A | N/A | N/A | N/A |
| 11. Shabani & Fisher (2006) [b c] | N/A | N/A | 1 | N/AN | 1 | N/A |
| 12. Taylor & Levin (1998) [a] | N/A | N/A | N/A | N/A | N/A | N/A |
| 13. Valentino et al. (2018) [c] | 2.436 | N/A | 0.43 | N/A | -0.14 | N/A |

[a] no baseline/all baseline points were zero (for at least one graph)
[b] only baseline phase was present (for at least one graph)
[c] only one participant data used
[d] desired decrease in independent variable occurred

**Table 3:** Mean Study Effect Sizes.

**Publication bias:** To address the "file drawer effect," the number of studies with results of zero required to reduce the overall effect to insignificant or suspect levels was determined for IRD and Tau-*U* effect sizes. Effect sizes for SMD were not used because of its tendency to inflate scores. It would take an additional 11 cases with an average effect size of 0 to bring the overall IRD into the ineffective range (<.36), and 16 cases with an average effect size of 0 to bring overall Tau-*U* into the small to ineffective range (<.20). Most SCRD studies contain between one and six participants. Using the average of three participants per study, at least 4 unpublished SCRD studies using IRD data and 5 for the Tau-*U* data would be needed to reduce currently obtained effect sizes to small or questionable levels.

## Discussion

The purpose of the current study was to conduct a meta-analysis of the literature on stimulus fading with children and adolescents. Effect sizes were calculated using standard mean difference (SMD), improvement rate difference (IRD), and Tau-*U*. The obtained effect values indicated that stimulus fading was a moderately effective treatment with some studies effect sizes in the high range. However, there is a threat of publication bias because the number of studies required to reduce effects to the small to questionable levels was quite low. In addition, several studies lacked data to calculate effect sizes either because of the absence of a baseline phase or including only a baseline phase but not subsequent intervention [9,11,16,41]. Effect sizes without an AB contrast cannot be calculated. Further, the reviewed studies were of moderate quality. However, six of the 13 studies were published before 2005 when Horner et al. [33] developed quality indicators for SCRD studies. Finally, no previous reviews have been conducted examining the effectiveness of stimulus fading interventions to improve children's discrimination learning in which to compare results to the current meta-analysis.

### Moderate Effectiveness Considerations

It was somewhat surprising that the body of literature reviewed was only moderately effective for improving children's discrimination learning. Most of the participants in the reviewed studies were children with autism or an autism spectrum disorder. Of the 13 studies, five were published in the journal *Research in Autism Spectrum Disorders* and the remaining eight were published in the *Journal of Applied Behavior Analysis*. These journals have extremely high quality for the process of peer review, although their impact factors are relatively low. However, it is important to note that impact factor is only a measure that reflects the average number of citations to articles published in journals, books, and dissertations [43]. Further, these two journals focus on very specific topics—with one publishing articles solely on autism, compared, for example, to the journal *Psychological Bulletin* which has a very high impact factor of 13.25. However, they publish only high quality systematic reviews from a plethora of scientific fields which would draw many more citations across disciplines.

Effect sizes do not always capture the nature of the SCRDs used to assess the effects of stimulus fading. SCRD effect sizes require an AB contrast (i.e., baseline data and intervention data), but not all stimulus fading studies use a baseline. Rather, they may begin with phase 1 of fading (e.g., very salient cue), move the phase 2 with a less salient cue, and phase 3 with a minor cue [42]. There would be no way to calculate effect sizes for the Taylor and Levin study, yet visual inspection of the data shows the process to have been very effective. The most striking example of an effective stimulus fading procedure that cannot be captured with effect sizes was conducted by Schreibman [16]. She trained six children with autism to make visual discriminations using forms drawn on black cards as stimuli. A finger prompt was gradually faded. Data were collected on the number of trials learned and not learned with a prompt and without a prompt. These data were displayed in a table and demonstrated the effectiveness of the procedure but did not lend themselves to typical SCRD graph display.

## Breadth of Stimulus Fading Techniques

Although not directly related to the effectiveness issue addressed previously, the corpus of literature on stimulus fading with children, mostly with autism, demonstrates its wide variety of applications and settings. In the latter case, stimulus fading techniques have been used in family homes, school settings including classrooms, dental and medical offices, hospitals, and outpatient clinics [6,10,11,12,17,39,41]. Stimulus fading has been used to address a variety of problematic behaviors and teaching discrimination skills for engaging in appropriate behaviors such as tolerating blood being drawn, correct tooth brushing, increasing the frequency of verbal interactions with peers and decreasing self-injurious behaviors [5,12,20,42]. These types of studies have been conducted with children for over 40 years [16] and continue to the present [17]. Hence demonstrating again that calculating effect sizes is not always the best method for assessing an interventions effectiveness—a point discussed in the next section.

## Limitations of Effect Sizes for Stimulus Fading Procedures

Almost all journals that publish systematic reviews require the authors to calculate effect sizes—and for good reason, especially when reviewing studies using traditional experimental designs consisting of control and treatment groups [44]. The typical approach using Cohen's *d* is calculated by subtracting the control group mean from the treatment group mean and dividing by the pooled standard deviation [45]. The more participants in both control and treatment groups, the greater the power and conclusions that can be drawn from obtained effect sizes [46].

This method for calculating effect sizes for SCRD studies is problematic because of small number of observations and floor effects limiting variability and results in overestimates of the parametric treatment effects [33,35]. Consequently, the most frequently used effect sizes are nonoverlap techniques [47]. The less overlap in data points between baseline (phase A) and intervention phase B) results in a higher effect size. Consequently, they compliment visual analysis of graphed data and are easy to calculate. However, they have two disadvantages that make them sometimes difficult to interpret: (a) monotonic (i.e., increasing) baseline trend and (b) large improvement in intervention trend which would be poorly captured by an index of level only (i.e., nonoverlap). The first problem, monotonic baseline trend, can be accounted for by the Tau-*U* nonoverlap method [36]. However, the second issue is more problematic. For example, baseline data of 6, 4, 8, 7, 5 and intervention data of 9, 11, 10, 9, 12 would have 100% nonoverlap, and so would data with the same baseline numbers but with intervention data of 34, 29, 31, 36, 35. Yet, the magnitude of change the second intervention produced was much greater than the first AB contrast data. Therefore, researchers conducting systematic reviews of a body of SCRD literature should not automatically calculate effect sizes but take into consideration the type of intervention and designs used, such as in the case of stimulus-fading addressed in the current review and use visual inspection as the primary form of analysis.

## Conclusion

Stimulus fading techniques may be considered effective for improving children's discrimination learning for over 40 years. They have been successfully used to improve discrimination skills for children primarily with autism for a variety of behaviors, situations, and settings. This conclusion was drawn with the help of effect sizes calculated but primarily through visual analysis of the data. The overreliance on visual analysis in this meta-analysis runs counter to current focus for greater rigor in SCRD studies by the Institute of Education Sciences [48] whose position is that the calculation of effect sizes should be standard practice. However, this position should be mitigated based on the type of SCRD intervention reviewed and nature of the design(s) used in the included studies. In the present review, the nature of stimulus fading techniques do not always lend themselves well to SCRD effect sizes. Consequently, there were several studies in the current review for which effect sizes could not be calculated because of designs unique to the analysis of stimulus fading procedures and for lack of robust data. However, this potential confound should not take away from the long-standing benefits to children and adolescents' discrimination learning of a variety of skills.

## References

1. Mueller MM, Palkovic CM, Maynard CS (2007) Errorless learning: Review and practical applications for teaching children with pervasive developmental disorders. Psychology in the Schools 44(7): 691-700.
2. Skinner BF (1938) The behavior or organisms: An experimental analysis. NY: Appleton-Century, Newyork.
3. Baer DM, Wolfe MM, Risley TR (1968) Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis 1(1): 91-97.
4. Hull CL (1950) Simple qualitative discrimination learning. Psychology Review 57(7): 303-313.
5. Bishop MR, Kenzer AL, Coffman CM, Tarbox CM, Tarbox J, et al. (2013) Using stimulus fading without escape extinction to increase compliance with toothbrushing in children with autism. Research in Autism Spectrum Disorders 7(6): 680-686.
6. Cuvo AJ, Godard A, Huckfeldt R, DeMattei R (2010) Training children with autism spectrum disorders to be compliant with an oral assessment. Research in Autism Spectrum Disorders 4(4): 681-696.
7. Schiff A, Tarbox J, Lanagan T, Farag P (2011) Establishing compliance with liquid medication administration in a child with autism. Journal of Applied Behavioral Analysis 44(2): 381-385.
8. Luscre DM, Center DB (1996) Procedures for reducing dental fear in children with autism. Journal of Autism and Developmental Disorders 26(5): 547-556.
9. Birkan B, Krantz PJ, McClannahan LE (2011) Teaching children with autism spectrum disorders to cooperate with injections. Research in Autism Spectrum Disorders 5(2): 941-948.
10. Cuvo AJ, Reagan AL, Ackerlund J, Huckfeldt R, Kelly C (2010) Training children with autism spectrum disorders to be compliant with a physical exam. Research in Autism Spectrum Disorders 4(2): 168-185.
11. Freeman KA, Piazza CC (1998) Combining stimulus fading, reinforcement, and extinction to treat food refusal. Journal of Applied Behavior Analysis 31(4): 691-694.
12. Shabani DB, Fisher WW (2006) Stimulus fading and differential reinforcement for the treatment of needle phobia in a youth with autism. Journal of Applied Behavior Analysis 39(4): 449-452.
13. Ghuman JK, Cataldo MD, Meck MH, Slifer KJ (2004) Behavioral training for pill swallowing difficulties in young children with autistic disorder. Journal of Child and Adolescent Psychopharmacology 14(4): 601-611.
14. Terrace HS (1963) Discrimination learning with and without "errors". Journal of the Experimental Analysis of Behavior 6(1): 1-27.
15. Terrace HS (1963) Errorless transfer of a discrimination across two continua. Journal of the Experimental Analysis of Behavior 6(2): 223-232.

16. Schreibman L (1975) Effects of within-stimulus and extra-stimulus prompting on discrimination learning in autistic children. Journal of Applied Behavior Analysis 8(1): 91-112.

17. Valentino AL, LeBlanc LA, Raetz PB (2018) Evaluation of stimulus intensity fading on reduction of rapid eating in a child with autism. Journal of Applied Behavior Analysis 51(1): 177-182.

18. Reitman D, Passeri C (2008) Use of stimulus fading and functional assessment to treat pill refusal with an 8-year-old boy diagnosed with ADHD. Clinical Case Studies 7(3): 224-237.

19. Holmbeck GN, Lavigne JV (1992) Combining self-modeling and stimulus fading in the treatment of an electively mute child. Psychotherapy: Theory, Research, Practice, Training 29(4): 661-667.

20. Pace GM, Iwata BA, Cowdery GE, Andree PJ, McIntyre T (1993) Stimulus (instructional) fading during extinction of self-injurious escape behavior. Journal of Applied Behavior Analysis 26(2): 205-212.

21. Vriend JL, Corkum PV, Moon EC, Smith IM (2011) Behavioral interventions for sleep problems in children with autism spectrum disorders: Current findings and future directions. Journal of Pediatric Psychology 36(9): 1017-1029.

22. Cunningham CE, Cataldo MF, Mallion C, Keyes JB (1983) A review and controlled single case evaluation of behavioral approaches to the management of elective mutism. Child and Family Behavior Therapy 5(4): 25-49.

23. Labbe EE, Williamson DA (1984) Behavioral treatment of elective mutism: A review of the literature. Clinical Psychology Review 4(3): 273-292.

24. Sanok RL, Ascione FR (1979) Behavioral interventions for childhood elective mutism: An evaluative review. Child Behavior Therapy 1(1): 49-68.

25. Council for Exceptional Children (2014) Council for Exceptional Children standards for evidence-based practices in special education. Exceptional Children 80(4): 504-511.

26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Medicine 6(7): e1000100.

27. Gage NA, Lewis TJ, Stichter JP (2012) Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. Behavioral Disorders 37(2): 55-77.

28. Losinski M, Maag JW, Katsiyannis A, Parks-Ennis RP (2014) Examining the effects and quality of interventions based on the assessment of contextual variables: A meta-analysis. Exceptional Children 80(4): 407-422.

29. Maggin DM, Briesch AM, Chafouleas SM (2013) An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. Remedial and Special Education 34(1): 44-58.

30. Common EA, Lane KL, Pustejovsky JE, Johnson AH, Johl LE (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.

31. Mitchell M (2002) Engauge Digitizer.

32. Nagler EM, Rindskopf DM, Shadish WR (2008) Analyzing data from small N designs using multilevel models: A procedural handbook. Washington, DC: U.S. Department of Education.

33. Horner RH, Swaminathan H, Sugai G, Smolkowski K (2012) Considerations for the systematic analysis of single-case research. Education and Treatment of Children 35(2): 269-290.

34. Busk PL, Serlin RC (1992) Meta-analysis for single-case research. In: Kratochwill TR, Levin JR (Eds.), Single-case research designs and analysis: New directions for psychology and education Hillsdale, NJ: Lawrence Erlbaum Associates. pp: 187-212.

35. Scruggs TE, Mastropieri MA (2012) PND at 25: Past, present, and future trends in summarizing single-subject research. Remedial and Special Education 34(1): 9-19.

36. Parker RI, Vannest KJ, Brown L (2009) The improvement difference for single-case research. Exceptional Children 75(2): 135-150.

37. Rosenthal R (1979) The "file drawer problem" and tolerance for null results. Psychological Bulletin 86(3): 638-641.

38. Rosenberg MS, Adams DC, Gurevitch J (2000) Metawin. Sunderland, MA: Sinauer.

39. Ducharme JM, Worling DE (1994) Behavior momentum and stimulus fading in the acquisition and maintenance of child compliance in the home. Journal of Applied Behavior Analysis 27(4): 639-647.

40. Sarkoff RA, Taylor BA, Poulson CL (2001) Teaching children with autism to engage in conversational exchanges: Script fading with embedded textual stimuli. Journal of Applied Behavior Analysis 34(1): 81-84.

41. Brown JL, Krantz PJ, McClannahan LE, Poulson CL (2008) Using script fading to promote natural environment stimulus control of verbal interactions among youths with autism. Research in Autism Spectrum Disorders 2: 480-497.

42. Taylor BA, Levin L (1998) Teaching a student with autism to make verbal initiations: Effects of a tactile prompt. Journal of Applied Behavior Analysis 31(4): 651-654.

43. Marks MS, Marsh M, Schroer TA, Stevens TH (2013) Misuse of journal impact factors in scientific assessment. Traffic 14: 611-612.

44. Hedges LV (1981) Distribution theory for Class's estimator of effect size and related estimators. Journal of Educational Statistics 6(2): 107-128.

45. Cohen J (1988) Statistical power for the behavioral sciences. 2nd (Edn.). Hillsdale, NJ: Erlbaum.

46. Kraemer HC, Blasey C (2016) How many subjects? Statistical power analysis in research. 2nd (Edn.), Thousand Oaks, CA: Sage.

47. Parker RI, Vannest KJ, Davis JL (2011) Effect size in single-case research: A review of nine nonoverlap techniques. Behavior Modification 35(4): 303-322.

48. Shadish WR, Hedges LV, Horner RH, Odom SL (2015) The role of between-case effect size in conducting, interpreting, and summarizing single-case research (NCER 2015-002). Washington, DC: National Center for Institute of Education Sciences, U.S. Department of Education.