Examining Generalization Performance in a Conditional Discrimination Task for Learners with Moderate to Severe Intellectual and Developmental Disabilities

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Mo Chen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Joe Reichle, Adviser

July 2016

Acknowledgements

I am grateful to my adviser, Dr. Joe Reichle, for guiding me so carefully, kindly, and patiently through this rigorous PhD program. I am also grateful to my committee members, Dr. Frank Symons, Dr. Jennifer McComas, Dr. Michael Rodriguez, and Dr. Benjamin Munson, for all of their invaluable advice, patience and support. This study would not have been possible without the support of the Down Syndrome Association of Minnesota, and the Gigi's playhouse. Thanks to Jolene Hyppa-Martin and Hannah Julien for their help with the initial recruitment of families. Thanks to Dr. LeAnne Johnson for her help with the assessment tools. Many thanks to Alyssa Merbler for her help with data collection and assessment administration. Special thanks to Dr. Breanne Byiers for her help with almost every step of the study, from recruitment, research implementation, to the tough writing process. The biggest thanks goes out to all of the children and their families who participated in the study.

Thanks to everyone who managed to help me out these years. All I can do is try my best to become a better person and be able to help more people in future. I'd also like to thank my parents and family in China for their constant encouragement as I embarked on my journey to America to earn a PhD at this world-renowned university. I am grateful to my dad Jin-fa Chen, my mom Mei-ying Chen, my younger brother Si-yuan Chen, and my god-mother Xiu-hua Zhang for their unconditional love for me. Finally, coming from a place where car ownership is not quite necessary, I want to acknowledge my 1998 Nissan Altima. Despite the numerous parking tickets (parking rules in the US can be confusing!), flat tires, breakdowns, dead batteries, spinning out twice on wintry roads, and even one trip in and out of the impound lot, my Nissan Altima got me safely to and from campus, to a variety of off-campus research and work sites, and even let me explore the Minnesota countryside. For that I am very grateful.

All I lost through these years is all I got. Thanks for the journey.

Dedication

This dissertation is dedicated to the children of the world who have moderate to profound developmental disabilities. May the contribution it makes benefit you someday.

Abstract

This study examined how differences between teaching stimuli in a conditional discrimination task impacted the discrimination/generalization outcomes among learners with intellectual and developmental disabilities (IDD). Three color sets were identified in a pilot study among 30 typically developing adults for the experiment. Subsequently five young children diagnosed with Down syndrome and moderate to severe intellectual disabilities participated in the experiment that employed an adapted alternating treatment single case experimental design. Prior to the intervention, participants' performance within each color set of nine stimuli was baselined to ensure room for improvement. During intervention, each participant received three intervention conditions in an alternated manner. The three experimental conditions included two conditional discrimination conditions, that differed only in the degree of color difference (maximal or minimal) between two teaching stimuli (i.e., S_A and S_B), and a simple discrimination condition that involved only one teaching stimulus (i.e., S_A). The assignment of the three color sets to the three intervention conditions was counterbalanced across five participants. As soon as the participants reached the mastery criterion on an intervention condition generalization testing was implemented involving all nine stimuli within the color set.

Results suggested that the simple discrimination was acquired most quickly. The comparison of the generalization gradients collected during baseline and generalization testing demonstrated that a simple discrimination was not less effective than the two conditional discriminations in enhancing the generalization performance subsequent to intervention. The minimal-difference conditional discrimination was relatively more effective in enhancing the generalization performance at least within one stimulus class than the maximal-difference one (and the simple discrimination). Limitations and implications of this study are discussed.

Table of Contents

List of Tables

List of Figures

CHAPTER I: INTRODUCTION

Stimulus generalization is the tendency for an organism to respond to novel stimuli differentially based on their similarity to the stimulus or stimuli involved in learning (Shepard, 1987; Wisniewski, Church, & Mercado III, 2009). Without stimulus generalization, functional use of a newly acquired skill would be compromised whenever the learner encountered a novel situation (Vervliet et al., 2011). The tendency for a learned response to generalize is seen across all species ranging from invertebrates to humans (Ghirlanda & Enquist, 2003; Thomas, 1993; Wisniewski et al., 2009). Given its adaptive value and ubiquity, generalization has been described as a fundamental associative process (Pavlov, 1927), psychology's first law (Shepard, 1987; Wisniewski et al., 2009), and has been widely explored in diverse disciplines (Howard, 2000). However, for evolutionary adaption there are some circumstances under which it is obviously necessary to inhibit the tendency for a newly acquired behavior to generalize (Kimble, 1961). A basic learning process that can assist in appropriately restricting the range of generalization is *discrimination* which emphasizes responding differently to two or more stimuli that vary on stimulus characteristic(s) (Kimble, 1961; Sutherland & Mackintosh, 1971).

Despite the prevalence and importance of generalization across all species, it can be a challenging problem for learners with intellectual and developmental disabilities (IDD). First, learners with IDD often have difficulty in generalizing learned skills. This may be due, at least in part, to their significant limitations in cognitive and adaptive skills (Albin & Horner, 1988; Johnston, Reichle, Feeley, & Jones, 2012; Stokes & Baer, 1977; Stokes & Osnes, 1989).

In addition to generalization challenges learners with IDD also frequently exhibit discrimination deficits. Together, generalization and discrimination deficits can lead to two types of generalization errors that have been described as under-generalization and over-generalization (Horner, Bellamy, & Colvin, 1984). *Under-generalization* happens when a learner "over - discriminates" the novel conditions as different from the acquisition condition and thus does not perform the behavior in the novel condition(s) where it should be used. *Over-generalization* happens when a learner does not discriminate sufficiently between the conditions where it is appropriate to use and the conditions where a newly acquired behavior should not be used. Thus, both under- and over-generalizations are undesirable outcomes. Maintaining the generalized performance of a new behavior over time has been rarely examined empirically; yet represents an important area of inquiry considering memory limitations often encountered among individuals with significant intellectual disabilities (Henry & MacLean, 2002; Jarrold, Baddeley, & Phillips, 2002).

Tactics have been proposed to facilitate the fine balance between desirable discrimination and desirable generalization performance. Since the publication of "An Implicit Technology of Generalization" (Stokes & Baer, 1977) generalization has received increasing attention. For instance, Stokes and Osnes (1989) summarized major generalization enhancing strategies into three categories that include: (a) exploiting current functional contingencies with an emphasis on the role of natural selection by the consequences of behavior, (b) training diversely with emphasis on the use of diverse learning examples, and (c) incorporating functional mediators such as relevant physical or social cues, rules, and language. In particular, the idea of teaching a range of examples have been expanded and refined into *general case instruction* as an instructional framework aimed at optimizing generalization performance (Chadsey-Rusch & Halle, 1992; Horner, Sprague, & Wilcox, 1982; Johnston et al., 2012; O'Neill & Reichle, 1993). General case instruction highlights the use of both positive teaching exemplar(s) (i.e., the S+ under which the targeted response should be emitted) and negative teaching exemplar(s) (i.e., the S- under which the targeted response should not be emitted or an alternative response should be emitted) to enhance the discrimination performance between the S+ and the S- and more importantly the generalization performance over the novel S+ and S- (Johnston et al., 2012; O'Neill & Reichle, 1993).

In spite of advances, our knowledge of generalization among learners with IDD is modest. First, generalization outcomes are often documented via a dichotomous approach in which generalization either occurs or doesn't occur (Osnes & Lieblein, 2003; Stokes & Osnes, 1989). Probes used to assess generalization are often determined by convenience as an adjunct procedure of the intervention instead of being quantified by specific stimulus dimensions (Stokes & Osnes, 1989). Generalization is more often examined in novel conditions under which the newly learned behavior should occur. By contrast, it is fairly infrequent that generalization has been examined by exposing a learner to novel conditions under which the newly learned behavior should not occur. Thus, potential over-generalization is much less often documented.

Experimental demonstrations of generalization effects have been somewhat limited. Among a group of systematic reviews addressing interventions for learners with intellectual disabilities (ID) that reported on generalization (Browder & Grasso, 1999; Chowdhury & Benson, 2011; Koyama & Wang, 2011; Morse & Schuster, 1996; Snell et al., 2010; Taft & Mason, 2011), none explored whether studies examined generalization during both baseline and intervention conditions. Thus many studies have failed to provide even an A-B design in evaluating generalization. Additionally, the generalization maintenance has rarely been empirically examined in spite of its theoretical significance (Falcomata & Wacker, 2013; Stokes & Osnes, 1989).

Finally, how potential intervention variables, such as the selection of teaching stimuli, intervention intensity, and reinforcement schedule, influence the appropriate and inappropriate generalization has not yet been sufficiently examined (Osnes & Lieblein, 2003). Several investigations addressing the functional relationship between intervention variables and generalization focused primarily on exploring whether generalized outcomes could be promoted using strategies proposed by Stokes and Baer (1977) and Stokes and Osnes (1989) (Osnes & Lieblein, 2003). Conversely, there has been a dearth of evidence addressing how systematic changes in certain intervention variables may influence generalization outcomes. For example, in the case of general case instruction, there is evidence supporting its effectiveness in enhancing discrimination and generalization performance by selecting and implementing both positive and negative teaching exemplars (Johnston et al., 2012; O'Neill & Reichle, 1993). However, little evidence has addressed how different arrangements of intervention parameters (e.g., including but not limited to; the number of S+ and S-, the difference between S+ and S-, the intervention sequence of S+ and S-, the intervention intensity, and the mastery criterion) during discrimination learning may influence the generalized outcomes.

Considering the available evidence addressing generalization among learners with IDD; a necessary next step requires a greater attention to experimental demonstrations of how variations in intervention procedures may impact generalization and generalization maintenance. The current study is aimed to examine the degree of generalization across a continuum of conditions by scrutinizing learners' performance via implementation of a generalization gradient in an examination of intervention variables that influence under- and over-generalization of a newly taught behavior for learners with IDD.

Describing Discrimination and Generalization

In developing effective evidence based interventions it is important to address both discrimination and generalization. With respect to discrimination, researchers and practitioners differentiate *simple* from *conditional discrimination*. In a simple discrimination, a learner is taught that a response will be reinforced in the presence of a specific stimulus (i.e., S+), but will not be reinforced in the presence of another stimulus (i.e., S-). Importantly, the positive (S+) and negative (S-) functions of stimuli remain the same from opportunity to opportunity.

In a conditional discrimination, a learner is taught that a specific response will be reinforced under one specific set of stimulus conditions and another response be reinforced under a different set of stimulus conditions. In a conditional discrimination the contingencies are reversed as a function of a particular discriminative stimulus (Andrews, Halford, & Boyce, 2012; Axe, 2008; Catania, 1998). Generally speaking, engaging in simple discriminations represents a prerequisite for conditional discriminations. As a higher-order discrimination, conditional discriminations are essential to the acquisition of numerous adaptive skills (Orr & Mast, 2014; Reichle & Wilkinson, 2012). Matching to sample techniques have been customarily applied to teaching a range of conditional discrimination skills such as literacy, math, and social communication skills (e.g., Doughty & Saunders, 2009; Hammond, Hirt, & Hall, 2012; Kuhn, Chirighin, & Zelenka, 2010; Sidman, 1971; Zaine, Domeniconi, & de Rose, 2014).

Simple Discrimination and Generalization

A considerable body of literature has documented the impact of simple discrimination on generalization performance for both nonhumans and humans (see three reviews by Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981; Purtle, 1973). Using a generalization gradient (to be further described in the forthcoming section), some consistent findings have been accumulated for both nonhumans and humans by comparing the generalization performance subsequent to a simple discrimination intervention and a single stimulus intervention. In a single stimulus intervention, participants were taught to respond to one specific stimulus (i.e., S+), which typically resulted in their responses occurring most often in the presence of the discriminative stimulus that was designated as the S+. However, in a simple discrimination intervention where participants were taught to respond to an S+ but simultaneously taught not to respond to an S-, response bias frequently occurred. That is, the participants had a greater tendency to respond to stimuli more similar to S+ and a lesser tendency to respond to stimuli increasingly similar to the S-. After specifically teaching a learner to refrain from responding to the S- in a simple discrimination, there was a propensity for less over-generalization for stimuli less similar to S+ as well as less under-generalization for stimuli more similar to S+. This, in turn, resulted in more desirable generalization and less undesirable generalization performance (Ghirlanda & Enquist, 2003).

Conditional Discrimination and Generalization

Only one review paper (Axe, 2008) has addressed conditional discrimination. This review was specific to intraverbal behavior and emphasized conditional discrimination learning with almost no reference to the generalization performance in conditional discrimination. Between 1960 and 2015 fewer than 30 refereed articles were published addressing the impact of conditional discrimination on generalization using a generalization gradient as means to examine a continuum of generalization. Only two of these investigations were specific to learners with IDD (Lane & Curran, 1963; Risley, 1964). This is in sharp contrast to the abundant evidence on the effects of simple discrimination on generalization via a generalization gradient (e.g., Bloomfield, 1967; Dickinson & Hedges, 1986; Dougherty & Lewis, 1991; Grusec, 1968; Marsh, 1972; Nicholson & Gray, 1971; Wheatley & Thomas, 1974; Wilkie, 1972; Wills & Mackintosh, 1998; also see reviews by Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981).

There are at least several potential contributors to the limited attention to conditional discrimination generalization among learners with IDD. First, conditional discrimination tasks can be challenging for learners with IDD (O'Donnell & Saunders, 2003; Saunders & Spradlin, 1989, 1990; Serna, Dube, & McIlvane, 1997). Thus, investigations addressing conditional discrimination for this population have largely focused on examining approaches to promote conditional discrimination acquisition, but rarely systematically examined generalization (Chadsey-Rusch & Halle, 1992; Johnston et al., 2012). Second, a matching to sample task is often utilized to teach conditional discrimination skills for learners with IDD. However, this research is mainly conducted within the framework of stimulus equivalence instead of being geared toward the examination of stimulus generalization (Fields, Reeve, Adams, Brown, & Verhave, 1997; Livesey & McLaren, 2009; Sidman, 1969, 2000). With such limited attention to the generalization aspect of conditional discriminations it is important to conduct systematic and in-depth explorations in this area.

Generalization Gradient: A Tool to Examine Generalization and Discrimination

Generalization gradient is a visual representation of the response strength (or probability of response) as a function of stimulus value (or stimuli of varying degrees of similarity to the teaching stimulus or stimuli) (Pierce & Cheney, 2008). It is a useful tool to visualize and analyze learners' generalization performance across a range of stimuli subsequent to intervention on certain stimulus or stimuli, through examination of the gradient parameters such as form (i.e., the shape of the gradient such as whether the distribution is symmetrical), slope (i.e., the rate of the change), peak (i.e., the maximum level of responding), and area (i.e., the coverage under the gradient along the stimulus continuum or a portion of the stimulus continuum) (Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981).

A generalization gradient provides a useful tool in more closely scrutinizing the phenomenon of generalization while at the same time considering the role of discrimination. A generalization gradient quantifies the degree of stimulus generalization, or more broadly, stimulus control ("any difference in responding in the presence of different stimuli," Catania, 1992, p. 372). At the same time it provides some detail in addressing the point at which a learner begins to view a new stimulus as being different from or discriminable from stimuli that he/she has been reinforced for selecting during intervention. Second, typically, when researchers utilize a generalization gradient, they select a specific stimulus dimension (e.g., the degree of brightness, the wavelength of color, the frequency of sounds) that can be easily quantified and experimentally manipulated (Ghirlanda & Enquist, 2003; Wisniewski et al., 2009). Consequently, generalization performance can be examined across a wider spread of stimuli, resulting in a more comprehensive picture of generalization performance. This, in turn, can help identify potential patterns of under- or over-generalization and avoid a more dichotomous determination of whether generalization either occurred or didn't occur. Finally, by applying a generalization gradient it becomes possible to analyze how variations of intervention procedures may influence a learner's generalization performance (e.g., Guttman & Kalish, 1956).

Effects that Differences Between Teaching Stimuli Have on Generalization of Conditional Discriminations

Among a range of intervention parameters involved in a conditional discrimination, the difference between teaching stimuli requires particular attention. How the degree of difference between teaching stimuli utilized during simple discrimination intervention impacts generalization has been well explored in basic research. A common finding has been that as the difference between S+ and S- decreases, the degree of response bias tends to get larger (Ghirlanda & Enquist, 2003; Horner, Albin, & Ralph, 1986). For instance, Hearst (1968) taught a group of pigeons with simple discrimination on the line-tilt dimension. They were taught in three conditions in which the difference between S+ and S- varied. Across all three conditions, they were taught to peck a white key with a vertical line on it (i.e., 0° of tilt as the S+), but not to peck a white key with a line of different degrees of tilt on it. For one condition, the S- was a line of 90° tilt; for a second condition the S- was a line of 60° tilt; and for the third condition the S- was a line of 30° tilt. Generalization testing was conducted in extinction with six different line orientations (30° apart) and a blank key. Results revealed that the simple discrimination condition involving the smallest difference between S+ and S- (i.e., a 30° difference) required the greatest number of teaching opportunities for pigeons to reach the mastery criterion. Furthermore, as the difference between S+ and S- decreased, the pigeons had a higher level of response to stimuli in the region of S+ and a lower level of response to stimuli in the region of S-. This type of response bias was most prominent and largest for the condition with the smallest difference between S+ and S-. From an instructional perspective, in a simple discrimination as S+ and S- get increasingly similar, there may be even less over-generalization on the side of S-. Additionally, there may be even less under-generalization to stimuli more similar to S+ (Ghirlanda & Enquist, 2003). This common finding applies to both human and nonhumans (Baron, 1973).

Despite potentially practical implications that can be derived from the simple discrimination literature; there is little evidence to support whether these implications apply to conditional discriminations. To date, only two studies (Eckerman, 1970; LaBerge, 1961) have used a generalization gradient in studying conditional discriminations. How differences between teaching stimuli in conditional discrimination tasks may impact the generalization performance in learners with IDD has been even rarely explored. In one exception, Horner and colleagues (1986) examined how the magnitude of the difference between teaching stimuli impacted learners' generalized use of conditionally produced rejecting skills. During grocery shopping they taught five adults with IDD to select 10 items that matched picture cards that represented the objects. Additionally the participants learned to reject a set of items that did not match picture cards that represented the objects in either of the following two conditions; (a) the items to be rejected were maximally different from those to be selected (i.e., a maximal-difference intervention condition), or (b) the items to be rejected were minimally different from those to be selected (i.e., a minimal-difference intervention condition). Intervention sessions to criterion for the two intervention conditions were reported to be equivalent and both intervention conditions were sufficient for the learners to correctly select the 10 trained items in a previously untrained store. However, a minimal-difference intervention condition was necessary for the learners to correctly reject items that had not been subject to intervention procedures in the store that had not been previously associated with intervention. Since the differences between teaching items were not clearly specified, it was not possible to quantify the differences along specific stimulus dimension(s) via a generalization gradient to gain a complete picture of learners' generalization performance, and explore whether generalization performance maintained. Finally, it was not possible to examine whether a small difference between teaching stimuli was necessary to prevent under- and over-generalization (while maintaining an efficient level of discrimination performance) for learners with IDD. Since Horner et al.'s (1986) investigation no studies have described generalization performance in conditional discrimination tasks for learners with IDD as a function of the difference between teaching stimuli. To refine our knowledge and further guide practical interventions, systematically applying a generalization gradient to examine more subtle aspects of discrimination and generalization performance represents a reasonable direction for research in this area.

Purpose of the Current Study and Research Questions

The literature describing and analyzing generalized outcomes in conditional discriminations for learners with IDD remains limited (Saunders & Spradlin, 1989; Sidman, 2000; Waston, 1966). No studies have examined generalization across a continuum of stimuli to determine generalization limits in the applied literature. This represents a substantial limitation in the discrimination and generalization literature. Being able to determine where breakdowns in these abilities occur has importance for a large number of functional skills.

The current study is aimed at examining the discrimination/generalization performance in a conditional discrimination task for learners with moderate to severe IDD using a generalization gradient. Of particular interest is whether the degree of differences between teaching stimuli (larger or smaller) impacts the discrimination/generalization performance. Specifically, the research questions addressed in the present study are:

1. Does a maximal-difference, or a minimal-difference (between two teaching stimuli S_A and S_B) conditional discrimination condition, or a simple

discrimination condition result in the fewest teaching opportunities to reach mastery criterion?

Hypothesis: The simple discrimination condition will require of the fewest teaching opportunities to reach mastery criterion, followed by a maximal-difference conditional discrimination condition, and then a minimal-difference conditional discrimination condition.

2. Does a maximal-difference, a minimal-difference (between two teaching stimuli S_A and S_B) conditional discrimination condition, or a simple discrimination condition best enhance the generalization performance?

   Hypothesis: The two conditional discrimination conditions will better enhance generalization performance than the simple discrimination condition. The minimal-difference conditional discrimination condition will better enhance the generalization performance than the maximal-difference conditional discrimination condition.

CHAPTER II: REVIEW OF LITERATURE

The ability to appropriately generalize skills across people, settings, stimuli, and across time is important for many aspects of daily life. Additionally, individuals with IDD frequently display discrimination/generalization deficits that may result in over-generalization and/or under-generalization. Some strategies to promote generalization were initially described by Stokes and Baer (1977). Since that article, there has been increased interest in the implementation of strategies to promote generalization among learners with IDD. In particular, using a range of teaching exemplars that include positive (stimulus conditions associated with use of the new behavior) and negative teaching exemplars (stimulus conditions that should not be associated with use of the new behavior), have been promoted to enhance generalization (Chadsey-Rusch & Halle, 1992; O'Neill & Reichle, 1993). Although a few examples of these approaches exist in the research literature (e.g., Albin & Horner, 1988; Day & Horner, 1986; Sprague & Horner, 1984), few studies have examined how these approaches may affect generalization among learners with IDD. Some key methodologies from basic learning literature may be relevant to this discussion.

In this chapter, the key concepts in relation to stimulus generalization are elaborated. Two approaches to documenting or measuring stimulus generalization are described. Subsequently, an overview of generalization literature relevant to learners with IDD is provided. This section will conclude with a critique of the existing literature. As a result of this critique, the remainder of this chapter will synthesize relevant studies with respect to: (a) evidence on stimulus generalization in simple discrimination tasks using the tool of a generalization gradient; (b) evidence on stimulus generalization in conditional discrimination; and (c) evidence on the generalization performance as a function of the difference between teaching stimuli in conditional discrimination.

Defining Generalization and Discrimination

In this section, the definitions of stimulus generalization as well as stimulus control are first explained followed by those related to discrimination.

Stimulus Generalization and Stimulus Control

There are two major types of generalization: stimulus generalization and response generalization. *Stimulus generalization* can be operationally defined as "the occurrence of relevant behavior under different, non-training conditions (i.e., across subjects, settings, people, behaviors, and/or time) without scheduling the same events in those conditions as had been scheduled in the training conditions" (Stokes & Baer, 1977; p. 350). *Response generalization* refers to the occurrence of a different behavior but serving the same function as previously taught behavior, while the stimulus remains unaltered (Duker, Didden, & Sigafoos, 2004).

Fundamentally, stimulus generalization can be conceptualized from the perspective of stimulus control. Broadly, the term *stimulus control* refers to "any difference in responding in the presence of different stimuli" (Catania, 1992, p. 372). Terrace (1966) provided the following more elaborated description of stimulus control:

> "Stimulus control refers to the extent to which the value of an antecedent stimulus determines the probability of occurrence of a conditioned response. It is measured as a change in response probability that results from a change in stimulus value. The greater the change in response probability, the greater the degree of stimulus control with respect to the continuum being studied." (p. 271)

Therefore, stimulus control reflects a functional relationship between the presentation of an antecedent stimulus and change in the probability of a response (i.e., antecedent stimulus →

response [R]); Terrace, 1966). In this sense, stimulus generalization occurs when a learner responds in the same way regardless of the change in antecedent stimuli.

Two types of antecedent stimuli include establishing operation (EO) and discriminative stimulus (S^D: S+ and S-). Specifically, the concept of *establishing operation* was initially termed by Keller and Schoenfeld (1950) but later refined by Michael (1982, 1993, 2000). Michael (1982, 1993, 2000) defined an EO as any stimulus, condition, or event that (a) momentarily alters the value of some stimulus as a reinforcer and (b) evokes all responses that have produced that reinforcer in the past. EO was later replaced with *motivating operation* (MO; Laraway, Snycerski, Michael, & Poling, 2003). This change occurred to encompass both the establishing and abolishing operations so as to more accurately represent the bidirectional effects of establishing and abolishing the value of reinforcement and the corresponding evoking or abating of the relevant behavior.

In contrast with EO or MO, a discriminative stimulus signals the availability of reinforcement (e.g., Tiger, Hanley, & Heal, 2006). In particular, a discriminative stimulus that is correlated with the availability of reinforcement contingent on a predetermined response is designated as S+. Other antecedent stimuli that do not signal an opportunity for reinforcement can be described as part of the discriminative stimulus correlated with the unavailability of reinforcement and are designated as S-. Although motivating operations and discriminative stimuli are different, both are able to exert stimulus control over a learner's responses (Pierce & Cheney, 2008).

Discrimination

Two types of discrimination learning include simple discrimination and conditional discrimination (Sutherland & Mackintosh, 1971).

Simple discrimination. A simple discrimination can be represented by a three-term contingency in which one response comes under the control of one stimulus that signals the availability of reinforcement (Ahearn, MacDonald, Graff, & Dube, 2007; Catania, 1998). A learner's response is reinforced in the presence of the specific stimulus but not reinforced in the absence of the stimulus or in the presence of another stimulus. Importantly, the positive (S+) and negative (S-) functions of stimuli remain the same from opportunity to opportunity. An example of simple discrimination would occur when a learner is presented with a red balloon and one or more balloons of other different colors. If the learner consistently selects the red balloon in the presence of balloons of other color(s) as a result of a reinforcement history in which selecting the red balloon has been reinforced but selecting the balloon of other color(s) has not been reinforced, he or she has demonstrated a simple discrimination. In this simple discrimination the S+ (i.e., red balloon) and the S- (i.e., balloon of other color[s]) do not change across discrimination opportunities.

Conditional discrimination. When compared with simple discrimination, a conditional discrimination is regarded as a higher-order discrimination. It can be described by a four-term contingency in which the three-term contingency (simple discrimination contingency) is altered by another stimulus (Sidman, 1986). That is, a response in the presence of a particular stimulus is followed by reinforcement only if another stimulus, the *conditional stimulus* or the *sample*, is present (Saunders & Spradlin, 1989). In conditional discrimination, the S+ and S- functions of stimuli change from opportunity to opportunity depending on which sample is present as an instructional cue for the learner. The relationship between selecting a stimulus and the reinforcing consequence is determined by another stimulus (a sample or instructional cue). More generally, in a conditional discrimination, (C_1: A→S^R, B→no S^R; C_2: A→ no S^R, B→ S^R), where C_1 and C_2 represent two different conditional stimuli or samples, A and B represent two different discriminative stimuli or two choices for responding, and S^R and no S^R indicate whether the responses result in reinforcement.

For example, given a red balloon sample and red and green balloon choices; selection of the red balloon results in the delivery of a positive reinforcer. In this example the learner would not receive reinforcement if he or she selected the green balloon choice. Alternatively, given the sample of a green balloon, the learner would be reinforced contingent on the selection of a green balloon choice in the presence of red and green balloon choices. Conditional discrimination plays an instrumental role in the acquisition of diverse functional skills including but not limited to literacy, math, and social communication skills (e.g., Doughty & Saunders, 2009; Hammond et al., 2012; Kuhn et al., 2010; Sidman, 1971; Zaine et al., 2014). Specific techniques for teaching conditional discrimination have been investigated extensively.

Matching to sample. Matching to sample paradigms have been commonly used in the extant literature to teach a wide range of conditional discriminations that include but are not limited to concept formation, reading, spelling, communication, and math (Carter & Werner, 1978; Cumming & Berryman, 1961; de Souza, Goyos, Silvares, & Saunders, 2007; Dube, McDonald, McIlvane, & Mackay, 1991; Dymond, Roche, Forsyth, Whelan, & Rhoden, 2008; Hammond et al., 2012; Murphy & Barnes-Holmes, 2010; Sidman, 1971; Stromer & Stromer, 1990, and a plethora of others). The preceding example with balloons is exemplary of a matching to sample procedure. Throughout an intervention session, the learner needs to make choices based on the specific sample presented for reinforcement (Fields, Garruto, & Watanabe, 2010). In the earlier example presented, the matching to sample procedure (referred to as *two-sample two-choice)* can be made more complex by adding a greater number of sequentially presented samples and/or simultaneously presented choices (Murphy & Barnes-Holmes, 2010).

Depending on the physical relationship between samples and choices, matching to sample procedures can be separated into identity and nonidentity matching (Reichle & Wilkinson, 2012). *Identity matching* means that a choice is identical to a sample. For instance, the learner uses a sample (e.g., a picture of a balloon) as an instructional cue and selects the exactly same picture from a set of choices. *Nonidentity matching* refers to a sample for which the choice is functionally or perceptually similar, but is not identical to the sample (Reichle & Wilkinson, 2012).

Depending on the temporal relationship between samples and choices, matching to sample tasks can be also categorized as *simultaneous* matching, and *non-simultaneous* or delayed matching. In *simultaneous matching*, the sample and choices can be viewed at the same time. In *non-simultaneous* or *delayed matching*, the sample and choices cannot be viewed at the same time. In other words, a delay temporally separates the disappearance of the sample and the appearance of the choices for a responding opportunity. Delayed matching is commonly used to study memory (Sargisson & White, 2001).

Finally, depending on the presentation of choices, matching to sample tasks can be grouped into *non-successive* matching and *successive* matching. In *non-successive matching*, at or after the presence of a sample, the choices are presented concurrently for the learner to make a selection for reinforcement. For a simple two-sample two-choice non-successive matching the contingency can be described as: $A_1 \rightarrow B_1 \mid B_2 \rightarrow$ RESPONSE TO B_1; $A_2 \rightarrow B_1 \mid B_2 \rightarrow$ RESPONSE TO B_2. For instance, a learner is presented with two fruit choices (i.e., an apple and an orange) concurrently, he or she is taught to select the apple at the presence of one sample (e.g., a picture of apple), but select the orange at the presence of the other sample (e.g., a picture of orange).

In *successive matching*, the choices are not presented concurrently for the learner to make a selection. Instead, at or after the presence of a sample, each of the choices is presented successively and the learner needs to decide whether a response should be emitted at the presence of a choice according to the sample presented. The contingency can be described as: $A_1 \rightarrow B_1 \rightarrow$ RESPONSE; $A_1 \rightarrow B_2 \rightarrow$ NO RESPONSE; $A_2 \rightarrow B_1 \rightarrow$ NO REPONSE; $A_2 \rightarrow B_2 \rightarrow$ RESPONSE. For instance, at the presence of a picture of apple as the sample, when an apple choice is presented, the learner is taught to select it. Alternatively, when an orange choice is presented, the learner is taught not to select it. However, given a picture of orange as the sample, when an apple choice is presented, the learner is taught to select it, and when an orange choice is presented, the learner is taught not to select it.

Overall, compared to non-successive matching, successive matching has been used sparingly in research. Its primary use has been in the study of short-term memory in the animal literature (Frank & Wasserman, 2005; Grant & Spetch, 1991; Nelson & Wasserman, 1978; Urcuioli, 2011; Urcuioli & Swisher, 2012; Wasserman, 1976). However, successive matching is more common in everyday life whereas non-successive matching is more of a standard matching to sample paradigm (Newhall, Burnham, & Clark, 1957). The main advantages of successive matching are, (a) the potential interactions between two or more sensory signals from the choices are minimized, (b) the responses due to position bias or preferences are also minimized, and (c) the likelihood of incorrect responses resulting from failing to remember the sample stimulus is also minimized (Wasserman, 1976).

In this section, stimulus generalization and stimulus control were defined, accompanied with an elaboration of two types of discrimination learning that would impact generalization performance. In the section to follow, how stimulus generalization has been usually reported or measured would be covered from a methodological viewpoint.

Measuring Stimulus Generalization

There are two main tactics used to document and measure stimulus generalization. In the applied behavioral research, a dichotomous approach is often used, while in the basic behavioral research, a generalization gradient is often used.

Dichotomous Approach

In the applied behavioral research, generalization is often examined by exposing the participants to the novel condition(s) and then documenting whether generalization occurs. Generally, researchers have recorded generalization outcomes in a dichotomous manner by whether the generalization occurs or does not occur. Broadly, two types of stimulus generalization errors consisting of both under- and over-generalization have been commonly categorized. Under-generalization is often reported when the target behavior did not occur in a novel condition where it should or could occur. For instance, a young learner with IDD might label a real dog as "dog" but not label a picture of dog as "dog." In contrast, over-generalization is generally reported when the target behavior occurs in a novel condition where it should or could not occur. For instance, a young child who just starts to speak calls every woman with long hair he or she runs into as "mom."

In particular, Horner and colleagues (1984) described three types of inaccurate stimulus control that could lead to under- and/or over-generalization, including the following: (a) *irrelevant stimuli controlling the target response*, meaning that the learner performs the response under stimulus conditions that contain the irrelevant stimuli, but does not perform the target response under stimulus conditions that do not contain the irrelevant stimuli, (b) *irrelevant stimuli controlling irrelevant responses*, meaning that the learner does not perform the target response under novel conditions because irrelevant stimuli in the novel conditions exert more powerful control over irrelevant responses than the relevant stimuli exert on the target response, and (c) *restricted stimulus control*, meaning that a response that should be under the control of multiple relevant stimuli or multiple characteristics of a relevant stimuli, but is only controlled by a subset of those stimuli.

In documenting generalization in applied behavioral research, a dichotomous approach has seldom been criticized in the extant literature. However, from a perspective of measurement, a dichotomous measure does not capture the performance differences across a continuum of slightly varying generalization probes. A more continuous approach to measurement would better allow pinpointing the point at which breakdowns occur.

A More Continuous Generalization Measure

Different from the dichotomous approach to quantifying generalization, the more continuous approach reflected in *generalization gradient* describes response strength (or probability of response) as a function of stimulus value (or stimuli of varying degrees of similarity to the teaching stimulus or stimuli; Pierce & Cheney, 2008). Thus, it is a tool to measure the limits of both discrimination and generalization, or stimulus control (Honig & Urcuioli, 1981). Compared to a more dichotomous approach, it is a much more sensitive measure for generalized and discriminated outcomes by capturing a more representative generalization and discrimination performance across a specific well-defined continuum of probes.

Generalization gradients can be described and compared on the basis of four indicators: form, slope, height, and area (Honig & Urcuioli, 1981). The concept of *form* or *shape* has not yet been precisely defined but can be examined from different aspects. Some hypothetical examples are provided in Figure 1. Across all these examples, the hypothetical dependent measure was the percent of responses in 10 opportunities. One way to characterize form is to examine whether the response distributions to either side of S+ on the test continuum are symmetrical (Honig & Urcuioli, 1981). In Figure 1(a), the response distribution to either side of 550 nm as the S+ is symmetrical, suggesting equivalent degrees of stimulus generalization or stimulus control on both sides of the S+. However, the response distribution is not symmetrical around the S+ in Figure 1(b), reflecting less stimulus generalization or a higher degree of stimulus control on the left side of the S+. Another way to describe form is to examine whether the gradient is monotonically increasing or decreasing across the test stimuli or along certain portion of the test stimuli. For instance, the gradient in Figure 1(c) is monotonically increasing across the continuum of sound intensity, and the gradient in Figure 1(d) is monotonically decreasing across the continuum of sound intensity. Similarly, in Figure 1(a) the gradient to the left side of the S+ is monotonically increasing across a portion of the spectral continuum and the gradient to the right side of the S+ is monotonically decreasing across a portion of the spectral continuum. Some other ways to describe the form may include but not limited to whether the gradient is in a peaked shape, and whether the gradient is bi-modal, or sigmoidal. Figures 1(a) and 1(b) are instances of a peaked gradient with only one mode. Figure 1(e) is an instance of a bi-modal gradient with two modes. Figure 1(f) is an instance of a sigmoidal gradient, reflecting generally a low degree of generalization on one side of the spectral continuum, but a generally high degree of generalization on the other side of the spectral continuum, with an intermediate degree of generalization for the spectral values in the middle of the continuum.

Slope is defined as the rate of change in responding between two points along the gradient, one of which is usually the teaching value. Slope is regarded as the most sensitive index of stimulus control or stimulus generalization, although lacking quantitative evidence (Honig & Urcuioli, 1981). Figure 2 presents two generalization gradients with different slopes. The relatively steeper gradient suggests less stimulus generalization and a higher degree of stimulus control, while the relatively shallower or flatter gradient suggests more stimulus generalization and a lower degree of stimulus control.

Height, also referred to as *peak*, is defined as the maximum level of responding along the gradient. It reflects the point(s) where the largest amount of generalization occurs. Height is usually related with the gradient slope (Honig & Urcuioli, 1981). For example, in Figure 2, although the peaks of both gradients are located at the stimulus of 550 nm, the height of the gradient with a larger slope is a level of 100% responses, while the height of the gradient with a smaller slope is around a level of 80% responses. In other words, in this case the largest amount of generalization occurs at the 550-nm stimulus for both gradients; however, the height of the steeper gradient is larger than that of the shallower one.

Finally, the *area* is defined by the total gradient, taking into account both the distribution of responses and the range of test values. The area under the gradient or a portion of the gradient can be mathematically computed. For instance, in Figure 1(b) the maximal area to either side of the S+ (550 nm) is five units, while the exact area to the right side of the S+ is 1.45 units, and the exact area to the left side of the S+ is 2.55 units. That is, there is a larger amount of generalization on the right side of the S+ compared to that on the left side of the S+. Therefore, to a certain extent, the indicator of area quantifies the amount of generalization across the test stimuli or a portion of the test stimuli.

To date there has not been consensus on how to use these indicators to quantify the degree of stimulus control and stimulus generalization. Usually, all or some of these indicators are used in combination to explain the generalization performance. In addition, for the most part, the two approaches to documenting and measuring stimulus generalization have not been integrated and applied together in either basic or applied research.

Evidence Addressing Stimulus Generalization for Learners with IDD

In this section, frequent challenges with stimulus generalization among learners with IDD will be described. Subsequently, intervention strategies aimed at addressing these challenges will be discussed (including a critique of the existing literature).

Challenges associated with stimulus generalization among learners with IDD have been documented for more than four decades (Scruggs & Mastropieri, 1984; Stokes & Baer, 1977). Baer, Wolf, and Risley (1968) argued that generalization should be "… programmed rather than expected or lamented" (p. 97). Stokes and Baer (1977) described the difficulties with generalization that were experienced by individuals with an unspecified degree of IDD who were “institutionalized” (p. 352); and suggested that problems with generalization may be the result of poor instructional strategies. Overall, these early researchers highlighted the issue of under-generalization in learners with IDD across a diversity of academic, social, and vocational skills.

One potential factor underlying the phenomenon of under-generalization for learners with IDD is stimulus overselectivity (Bailey, 1981; Lovaas, Schreibman, Koegel, & Rehm, 1971). Lovaas et al. (1971) defined overselectivity as responding to one, or to some reduced number, of the relevant components in a complex stimulus. For instance, when a learner labels a real dog as “dog” but does not label a picture of dog as “dog,” the labeling behavior is considered to be under the control of a restricted and possibly irrelevant stimulus. Thus, overselectivity may result in under-generalization as well as over-generalization due to its inaccurate or limited stimulus control. Lovaas and colleagues (1971) taught children with autism spectrum disorder (ASD), children with moderate to severe IDD but not ASD, and typically developing children to respond to a multidimensional stimulus containing auditory, visual, and tactual components. Following instruction, the participants were tested on the individual stimulus components to assess which cue or cues had become functional in controlling their behavior. Children with ASD responded primarily to only one of the three components (i.e., they were overselective). The typically developing children responded similarly to all three components. Children with moderate to severe IDD who did not have ASD performed somewhere in between the highly overselective children with ASD and the typically developing children who did not, for the most part, display stimulus overselectivity.

Although much of the research examining stimulus overselectivity has focused on learners with ASD, a number of studies have demonstrated overselectivity among learners with moderate to severe IDD (Dickson, Deutsch, Wang, & Dube, 2006; Dickson, Wang, Lombard, & Dube, 2006; Dube et al., 1999; Wilhelm & Lovaas, 1976). These studies suggested that the degree of overselectivity is directly related to the degree of intellectual disability. Although the specific stimuli that exert stimulus control in cases of overselectivity can be predicted and modified by manipulating reinforcement contingencies, it still remains unclear how to best reduce or prevent overselectivity among learners with moderate to severe IDD (Dube & McIlvane, 1997).

Finally, learners with IDD may have more challenges in maintaining their generalization performance due to memory limitations (Belmont & Butterfield, 1969; Henry & MacLean, 2002; Jarrold et al., 2002). Learners with IDD usually learn and retain new information less efficiently than age-matched typically developing individuals (Jarrold et al., 2002). Although lacking of evidence on this aspect, it is plausible that this group of individuals may have greater difficulty in maintaining not only the acquired skills but also the generalized performance due to the impairment in different subcomponents of memory, such as working memory (Alloway, 2010; Pickering & Gathercole, 2004; Schuchardt, Gebhardt, & Mäehler, 2010), short-term memory (Jarrold & Baddeley, 1997; Jarrold et al., 2002; Marcell & Armstrong, 1982; Marcell & Weeks, 1988), and long-term memory (Carlesimo, Marotta, & Vicari, 1997; Jarrold, Phillips, & Baddeley, 2007). Even if they have no challenges in maintaining their generalized outcomes, little is known regarding how their generalization performance may change over time.

Intervention Strategies for Promoting Stimulus Generalization

Intervention strategies aimed at improving stimulus generalization for learners with IDD have been explored by a number of investigators (Chadsey-Rusch & Halle, 1992; Johnston et al., 2012; O'Neill & Reichle, 1993; Stokes & Baer, 1977; Stokes & Osnes, 1989). Stokes and Baer (1977) reviewed strategies and summarized nine strategies reported by researchers to monitor or facilitate generalization. The strategies included: (a) Train and Hope, (b) Sequential Modification, (c) Introduce to Natural Maintaining Contingencies, (d) Train Sufficient Exemplars, (e) Train Loosely, (f) Use Indiscriminable Contingencies, (g) Program Common Stimuli, (h) Mediate Generalization, and (i) Train "To Generalize." Table 1 lists more detailed definitions for each of these strategies. They concluded that only some of these strategies were promising and expressed the concern that problems with generalization may be attributed to poor instructional programming (e.g., the use of "Train and Hope," "Sequential Modification," and/or "Introduce to Natural Maintaining Contingencies" strategies; Stokes & Osnes, 1989).

Stokes and Osnes (1989) further refined the work by Stokes and Baer (1977) and organized the general principles of generalization programming into three areas with several tactics under each area. These areas included: (a) *Exploiting current functional contingencies* with specific tactics that included contacting natural consequences, recruiting natural consequences, modifying maladaptive consequences, and reinforcing occurrences of generalization, (b) *Training diversely* with specific tactics that included using sufficient stimulus exemplars, using sufficient response exemplars, making antecedents less discriminable, and making consequences less discriminable, and (c) *Incorporating functional mediators* with specific tactics that included incorporating common salient physical stimuli, incorporating common salient social stimuli, incorporating self-mediated physical stimuli, and incorporating self-mediated verbal and covert stimuli.

Since Stokes and Baer (1977) and Stokes and Osnes (1989), there have been efforts to enhance the generalization performance in learners with IDD. There are a plethora of systematic literature reviews on diverse interventions for this group of population (Browder & Grasso, 1999; Carter, Sisco, Chung, & Stanton-Chapman, 2010; Chowdhury & Benson, 2011; Didden, Korzilius, van Oorsouw, & Sturmey, 2006; Hughes et al., 2012; Koyama & Wang, 2011; Morse & Schuster, 1996; Snell et al., 2010; Taft & Mason, 2011). Among studies addressed in these reviews, generalization that was reported ranged from 34% to as high as 86%. Many reviews also coded the reporting status of maintenance (or the generalization across time of the acquired skills under intervention condition[s]), although none of the review studies addressed the maintenance of generalization performance itself.

Several studies directly examined the degree to which the generalization tactics had been addressed in the existing literature. For instance, through examining a sample of studies published in three behavior analysis journals from 1990 to 2002, Osnes and Lieblein (2003) reported that researchers devoted increasing attention to arranging variables during acquisition to promote generalization, instead of relying on implicit "Train and Hope" technique. Specifically, forty-three percent of the 88 studies reviewed (*n* = 38) contained the terms "generalization" or "maintenance" in their titles, descriptors, and/or abstracts. Thirty of these studies reported generalization and eight reported maintenance. In addition, a majority of the 30 studies that reported generalization provided some overt generalization strategies in the procedures, such as exploiting current functional contingencies, and training diversely. Similarly, Hughes et al. (2012) reviewed studies on social interaction skill intervention across a range of secondary school students with ASD and/or IDD, published from 1980 through 2011. Among the 13 studies identified, eight reported strategies introduced to program generalization as described by Stokes and Baer (1977), including (a) training sufficient exemplars, (b) programming common stimuli, and (c) mediating generalization.

Falcomata and Wacker (2013) were interested in identifying whether tactics described by Stokes and Osnes (1989) were applied to promoting generalization in functional communication training (FCT). They conducted a systematic literature review focusing on FCT that involved an analysis of generalization across contexts, stimuli, or individuals in the epoch from 1985 to 2012. Among the 23 studies examined, they reported that only eight specified tactics for promoting generalization within the framework outlined by Stokes and Osnes (1989), and for the most part the tactics involved recruiting natural consequences, modifying maladaptive consequences, using sufficient stimulus exemplars, and programming common physical stimuli. In this sense, they concluded that only a small number of FCT studies had systematically evaluated generalization.

Among the different generalization enhancing tactics, “training sufficient stimulus exemplars” has become one of the widely applied to promote generalization. In particular, with the importance of discrimination learning on generalization performance taken into account, training sufficient exemplars has been refined into a more systematic approach of concurrently considering discrimination and generalization, which is referred to as *general case instruction*.

General Case Instruction

General case instruction has been characterized as an instructional framework that promotes appropriate discrimination and generalization through careful establishment of stimulus control over both positive (S+) and negative (S-) exemplars (e.g., Chadsey-Rusch & Halle, 1992; Horner et al., 1982; O’Neill & Reichle, 1993). A positive exemplar refers to a condition where a learner is taught to emit a target behavior under stimulus characteristics that should be associated with a behavior’s emission, while a negative exemplar refers to stimulus features where a learner is taught not to emit the target behavior, or emit an alternative behavior. O’Neill and Reichle (1993) delineated a procedure for general case instruction that consisted of six general steps that included; (a) determine the range of stimulus conditions in which it is appropriate to emit the target response, as well as the range of stimulus conditions in which it is not appropriate to emit the target response, (b) determine the range of relevant response variation and stimulus variation within each of the two conditions, (c) select positive teaching exemplars (S+) and negative teaching exemplars (S-) for use in teaching and probe testing, (d) sequence the teaching exemplars, (e) teach the exemplars,, and (f) test for acquisition with non-trained probe exemplars.

Stimulus control plays an important role in general case instruction (Horner et al., 1982). In general case programming, stimulus control is designed to occur in the presence of a range of stimuli that constitute a stimulus class. Within a stimulus class all stimuli share a common and specific set of characteristics that distinguish the stimulus class from other classes (Pierce & Cheney, 2008). When any member from a stimulus class is able to produce the same effect as a single stimulus (i.e., the S+), and stimuli that are *not* members of the stimulus class (i.e., the S-) do *not* produce the target behavior, then the general case has been learned (Chadsey-Rusch & Halle, 1992; Horner et al., 1982). Therefore, general case instruction emphasizes the concurrent implementation of both multiple positive and negative exemplars to produce well-differentiated responses between the two types of exemplars. It also promotes the generalization of the learned skills over other untrained positive and negative exemplars, that is, minimizing the under- and over-generalization errors.

General case instruction for learners with IDD. General case instruction has been utilized to teach learners with IDD many different functional skills that include dressing skills (Day & Horner, 1986), personal hygiene (Stokes, Cameron, Dorsey, & Fleming, 2004), street crossing (Horner, Jones, & Williams, 1985), vending machine use (Sprague & Horner, 1984), telephone use (Horner, Williams, & Steveley, 1987), fast food restaurant skills (Steere, Strauch, Powell, & Butterworth, 1990), photography skills (Giangreco, 1983), janitorial and housekeeping job tasks (Woolcock, Lyon, & Woolcock, 1987), communication skills (Albin & Horner, 1988; Chadsey-Rusch, Drasgow, Reinoehl, Halle, & Collet-Klingenberg, 1993; Kreibich, Chen, & Reichle, 2015), and academic skills (Chezan, Drasgow, & Marshall, 2012). A majority of these studies have demonstrated that general case instruction was effective in producing generalized effects (e.g., Chadsey-Rusch et al., 1993; Engelmann & Carnine, 1982; Horner et al., 1985; Sprague & Horner, 1984).

For instance, Sprague and Horner (1984) taught six high school students with moderate to severe IDD to use vending machines. Intervention was conducted in a multiple-baseline design across subjects to compare the effectiveness of three strategies that involved the use of; (a) a single vending machine, (b) three similar machines, and (c) three machines that sampled the range of stimulus and response variations in a defined class of vending machines. Results indicated that the third approach was most effective in achieving generalized performance on 10 non-trained vending machines. Chadsey-Rusch et al. (1993) examined the effects of general case instruction on requests for assistance by three adolescents with severe IDD. They compared the effects of single instance and general case instruction in a multiple-probe design across participants. During the single instance instruction, the learners were taught to request help in only one stimulus condition (i.e., requesting food). During the general case instruction, the stimulus conditions the learners were exposed to varied. Performance was assessed on nine generalization probes that occurred during natural opportunities to request help. Results revealed that two of the three learners required fewer intervention opportunities to reach the criterion for generalized outcomes during general case than during single instance instruction. Recently, Kreibich et al. (2015) taught a learner with ASD and IDD to request breaks from tasks while concurrently increasing task engagement via a tolerance for delay in reinforcement delivery procedure. In a multiple-probe design across tasks, the intervention was concurrently initiated with two tasks involving short periods of engagement prior to the learner's disengagement. Results showed that although generalization of break requests to untrained short engagement tasks did not occur until the intervention was implemented, overgeneralized use of break requests with long engagement tasks did not occur as well. In general, the existing evidence supports the utility of general case instruction in promoting generalization performance, with respect to minimizing both under- and over-generalization errors.

Critique of Generalization Literature Addressing Learners with IDD

In spite of the utility of general case instruction as well as other generalization intervention strategies, some potential limitations exist in the literature addressing the evidenced based applications for learners with IDD. First, generalization tends to be examined under conditions where the target skill should occur. Consequently, it is far less likely to be examined under the conditions where it should not occur. For example, among studies that focused on teaching learners with IDD to request or reject conditionally that were published between 1985 and 2015, only several studies (e.g., Reichle & Johnston, 1999; Kreibich et al., 2015) not only probed generalization for the communicative skill when it should occur to see potential under-generalization but also concurrently looked at the potential over-generalization of the communicative skill when it should not occur. As suggested by Kuhn (1962), progress in science and psychology not only derives from reviews of success, but from systematic documentation and analysis of failure. Without a complete exploration of potential generalization errors including both under- and over-generalization, it would be difficult for us to advance further toward the understanding of the mechanisms underlying generalization.

Rarely have investigators implemented a more continuous measurement across more subtle changes in stimulus conditions that a measurement tool such as a generalization gradient would permit. Not being able to observe the influence that more subtle stimulus differences may have on learner discrimination and generalization limits an examination of precise generalization challenges that may exist. One exception is the study by Reichle and Johnston (1999) who attempted to apply a generalization gradient while teaching two elementary school learners with severe IDD to conditionally request preferred snacks. The distance between the snacks and the learner served as the discriminative stimulus to signal when to emit the communicative request. Subsequent to discrimination intervention on the positive exemplar (i.e., snacks placed about 3 inches from the teacher and a request was appropriate) and the negative exemplar (i.e., snacks placed about 6 inches from the learner and a request was inappropriate), generalization testing was implemented consisting of four untrained distances (i.e., 5, 12, 16, 20 inches away from the learner). Although generalization testing did not include the teaching stimulus values and learners' responses on all generalization testing values were not displayed graphically, their study suggested the promising utility of a generalization gradient in applied research.

Currently, researchers have been slow to generate experimental demonstrations of generalization effects. Among the evidence on teaching learners with IDD to request or reject conditionally, only several studies (e.g., Horner, Albin, & Ralph, 1986; Kreibich et al., 2015; Sigafoos & Roberts-Pennell, 1999) have reported at least one generalization effect during both baseline and intervention conditions. Across nine systematic reviews addressing interventions for learners with IDD (Browder & Grasso, 1999; Carter et al., 2010; Chowdhury & Benson, 2011; Didden et al., 2006; Hughes et al., 2012; Koyama & Wang, 2011; Morse & Schuster, 1996; Snell et al., 2010; Taft & Mason, 2011) searched with "intellectual and developmental disability" and "review" as the keywords, none reported whether studies reviewed examined generalization by obtaining data during both baseline and intervention conditions.

Although maintenance of acquisition performance has been frequently discussed in the intervention literature addressing learners with IDD, the maintenance of generalized behavior has rarely been the focus of empirical investigations (Falcomata & Wacker, 2013; Stokes & Osnes, 1989). Broadly, the importance of documenting and exploring the maintenance of the acquired skills has been emphasized early on from the perspective of the generalization across time (Stokes & Baer, 1977; Stokes & Osnes, 1989), which has sometimes been referred to as "resistance to extinction," and "persistence" (Osnes & Lieblein, 2003). Nonetheless, the issue of the maintained generalization has primarily been discussed and investigated in basic behavior research involving animals (e.g., Hinson & Malone, 1980; Howard, 1979; Malone & Staddon, 1973; Schuster & Gross, 1969; Tennison & Hinson, 1993; White & Thomas, 1979; Winograd, 1965). In contrast, little empirical evidence in applied behavioral research has addressed the maintenance of the generalization outcomes, and little is known whether and how the generalized outcomes at one time point might change over time for humans including learners with IDD.

Finally, existing evidence has focused primarily on applying strategies to enhance generalization performance and much less on how functional variables involved in the strategies may impact learners' generalization outcomes. For example, consider the selection of positive exemplars and negative exemplars for general case instruction. As the positive and negative exemplars become increasingly different, it may be easier for a learner to discriminate between them. However, to date, it remains somewhat unclear how the spacing of these teaching examples may impact the learner's performance during acquisition or subsequently in efforts to scrutinize generalization. Similarly, it is unclear how other relevant factors such as the number of positive and negative exemplars may influence learners' discrimination efficiency and generalization performance in general case instruction.

In the following sections, an overview of studies that addressed; (a) the effects of simple discrimination on generalization performance using a generalization gradient as a dependent measure, (b) the effects of conditional discrimination on generalization performance, and (c) the effects of the difference between teaching stimuli in a conditional discrimination task on generalization performance, will be provided.

Studies Addressing Generalization in Simple Discrimination Using Generalization Gradient to Quantify Generalization Extent

In order to establish stimulus generalization as a research area in its own right and explore the relationship between generalization and the discriminability of stimuli, Guttman and Kalish (1956) examined stimulus generalization in pigeons. They reinforced pigeons for pecking a key in the presence of a particular wavelength of light on a variable interval reinforcement schedule. Each of four different groups of pigeons was taught to respond to unique spectral stimulus (530, 550, 580, 600 nm, respectively). After a steady rate of responding emerged, the pigeons were tested during extinction with a random and repeated presentation of 11 different spectral values that were evenly spaced around and included the teaching value. As Figure 3 shows, plots of responses as a function of spectral wavelength yielded orderly generalization gradients, with the highest rate of responding or the peak occurring around the teaching value and increasingly lower rates of responding at increasingly different test values. In addition, a steeper generalization gradient (e.g., the gradient with 600 nm as the teaching stimulus in this case) suggests stronger response strength to some stimuli than others, which may reflect a less amount of generalized responses. In contrast, a shallower one (e.g., the gradient with 530 nm as the teaching stimulus in this case) suggests relatively similar response strength to most stimuli, which may reflect a larger amount of generalized responses.

Interdimensional and Intra-dimensional Simple Discrimination

Subsequent to the study by Guttman and Kalish (1956) in which only one single teaching stimulus was involved, researchers started to investigate how a generalization gradient looks after simple discrimination intervention, in which discrimination between two stimuli was taught (Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981; Purtle, 1973). Typically, intervention has been conducted in a "go/no-go" experimental paradigm (Freeman, Kramarcy, & Lee, 1973; Rodgers & Thomas, 1982). That is, responding was reinforced in the presence of the positive stimulus (S+) but not reinforced in the presence of the negative stimulus (S-), with the S+ and S- unchanged.

According to the attributes of the stimuli involved in simple discrimination intervention, the intervention can be *interdimensional* or *intra-dimensional*. In the *interdimensional* simple discrimination intervention, the teaching stimuli differ in many stimulus dimensions and thus cannot lie along the same stimulus dimension. For instance, the S+ is a pure tone (under which the learner should press the button) and the S- is a yellow light (under which the learner should not press the button). Instead, in the *intra-dimensional* simple discrimination intervention, the teaching stimuli lie along the same stimulus dimension on which generalization tests are conducted (Ghirlanda & Enquist, 2003). For instance, the S+ is a 500-Hz tone, and the S- is a 700-Hz tone. More practically, in an intra-dimensional simple discrimination intervention, the S+ is a red light under which a pedestrian should walk across the street and the S- is a green light under which a pedestrian should not walk across the street. In an interdimensional simple discrimination intervention, the S+ is a written word "OPEN" in front of a coffee shop under which a person could walk into the shop, the S- is a written word "CLOSED" under which a person should not walk into the shop.

Compared to single stimulus intervention (e.g., Guttman & Kalish, 1956), simple discrimination, especially intra-dimensional simple discrimination intervention, would sharpen the generalization into a steeper or narrower shape, indicating a higher degree of stimulus control and a lower degree of generalization. For instance, Jenkins and Harrison (1960) conducted tonal frequency generalization tests in pigeons following; (a) non-differential intervention (i.e., a single stimulus intervention) in the presence of a 1000-Hz tone, (b) interdimensional simple discrimination intervention, in which the S+ was a 1000-Hz tone and the S- was silence, or (c) intra-dimensional simple discrimination intervention, in which the S+ was a 1000-Hz tone and the S- was a 950-Hz tone. The generalization gradient was steepest for pigeons under intra-dimensional simple discrimination intervention and flattest for pigeons under nondifferential intervention.

Peak shift. A substantial base of evidence demonstrates that intra-dimensional simple discrimination intervention not only results in a generalization gradient of a steeper shape but also shifts the peak of responding toward a value away from the S+ in the opposite direction of the S- (i.e., a *peak shift* effect; Akins, Drew Gouvier, & Lyons, 1981; Bloomfield, 1967; Cheng, Spetch, & Johnson, 1997; Dickinson & Hedges, 1986; Dougherty & Lewis, 1991; Galizio, 1985; Grusec, 1968; Lyons, Klipec, & Steinsultz, 1973; Marsh, 1972; Nicholson & Gray, 1971; Wheatley & Thomas, 1974; Wilkie, 1972; Wills & Mackintosh, 1998; and a plethora of others). In other words, the peak-shift effect implies that after an intra-dimensional simple discrimination intervention, a learner's maximum responding tends to be displaced from the S+ but located at another stimulus value on the side of the S+. Hanson (1959) gave four groups of pigeons intra-dimensional simple discrimination intervention on the wavelength dimension. Each group received the same S+ (i.e., a light of 550 nm) but different examples of the S- for the four groups (i.e., 555, 560, 570, or 590 nm, respectively for each group). The fifth group served as the control group and received non-differential intervention with the S+ of 550 nm only. After intervention, generalization tests consisting of 13 stimuli ranging from 480 to 620 nm (including the S+ and/or S- during intervention) were conducted in extinction. Only the control group showed a peak of responding at the S+. The four experimental groups showed a peak of responding to wavelengths below the S+ (i.e., a displacement away from the S- in the direction of the S+), with the size of the peak shift functioning inversely with the magnitude of the difference between the S+ and S-.

The majority of subsequent research has supported the generality of the peak-shift effect after intra-dimensional simple discrimination intervention. Specifically, the effect has been extended to various species including fish, rats, chickens, horses, birds, and humans. These studies have involved a variety of stimulus dimensions, such as light spectra, sound frequency, object orientation, object location, object size, and even relative numerosity (see Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981; Purtle, 1973, for reviews). In particular, peak-shift effect can even be invoked to explain the preference biases observed among many species (Derenne, 2010; Derenne, Breitstein, & Cicha, 2008). For example, a man who had a positive relationship with a woman with dark brown hair (S+) but a negative relationship with a woman with light brown hair (S-) would prefer women with very dark hair (Derenne, 2010; Powell, Symbaluk, & McDonald, 2002). Similarly, a man who had a positive relationship with an extrovert woman but a negative relationship with an introvert woman would prefer very extroverted woman (Derenne, 2010; Powell, Symbaluk, & Honey, 2009).

Even in cases in which a peak shift does not occur, intra-dimensional simple discrimination intervention typically results in an asymmetrical generalization gradient in which animals are more apt to respond to values on the S+ side of the distribution than to values on the S- side, an effect referred to as an *area shift* (Rilling, 1977). Peak shift has also been referred to as *response bias* (Ghirlanda & Enquist, 2003), meaning that there is/are a stimulus or stimuli that elicit/s stronger responding than S+. This response bias is often found subsequent to a simple discrimination between two stimuli differing along a certain stimulus dimension (i.e., intra-dimensional intervention). The shape of a generalization gradient has been shown to be influenced by two stimulus dimensions; rearrangement or intensity (Ghirlanda & Enquist, 2003). Stimuli along the rearrangement dimensions involve a rearrangement of stimulation with respect to S+, without significant change in the total activation of sense organs. For instance, different orientations of lines (e.g., 0°, 30°, 60°, 90°) in the visual field produce the same amount of stimulation in the eye, and thus the object orientation can be considered as a rearrangement dimension. Similarly, monochromatic light (as described previously) can also be considered as a rearrangement dimension because the total receptor activation is considered as almost constant over a large range of wavelengths across diverse species (Ghirlanda & Enquist, 2003). Conversely, stimuli along an intensity dimension are considered to involve changes in the total activation of sense organs (e.g., when the stimuli vary on sound or light intensity). The generalization gradient along the rearrangement dimension typically peaks at S+ after establishing an interdimensional simple discrimination, but peaks at a stimulus that is further away from S- than S+ (i.e., a peak shift or a response bias) after successful intra-dimensional simple discrimination intervention (see Figure 4 for an example). However, the generalization gradient along the intensity dimension is often monotonic (rather than peaked) over large ranges of intensity regardless of inter or intra-dimensional simple discrimination intervention, and typically shows larger response biases than rearrangement gradients (see Figure 5 for an example). Additionally, when the stimulus dimension involves changes of both the rearrangement and intensity, the generalization gradients yielded have been either peaked or monotonic, with response biases found along all dimensions (Ghirlanda & Enquist, 2003). Ghirlanda and Enquist

(2003) summarized the major stimulus dimensions according to the contribution from the rearrangement and/or intensity.

Rearrangement stimuli that typically produce generalization gradients in a peaked shape include object orientation, object location, object shape, tone frequency, and monochromatic light. Typically, rearrangement gradients peak at or near at the S+ after successful non-differential intervention or interdimensional simple discrimination intervention. However, they produce a peak shift subsequent to successful intra-dimensional simple discrimination intervention. Further, Ghirlanda and Enquist (2003) pointed out that a closer proximity between S+ and S- usually produced a gradient whose peak is higher and further away from S+. This finding supports Hanson's (1959) results that the magnitude of the peak-shift effect increases as the spacing between the S+ and S- decreases.

The potential implication of the findings from basic research is that the closer S+ and S- are, the discrimination of what is an S+ and S- is narrower. Moreover, there is a greater tendency to respond to stimuli on the side of S+ (or stimuli more similar to S+) and a lower tendency to respond to stimuli on the side of S- (or stimuli less similar to S+). A more practical implication may be that as S+ and S- get closer, there may be less over-generalization on the side of S- (or stimuli less similar to S+) as well as less under-generalization on the side of S+ (or stimuli more similar to S+). Ghirlanda and Enquist (2003) also mentioned that in addition to the factor of similarity or difference between S+ and S-, other intervention and testing procedures as well as the characteristics of stimuli used may impact the degree of peak shift or response biases. Given the potential empirical implications of the peak-shift effect for further understanding of generalization, it may be helpful to have a brief review of the existing theoretical accounts for the peak-shift effect in the section to follow.

Theoretical explanations for peak shift. Two major accounts pertaining to peak shift include Spence's (1937) explanation (in terms of summation of excitatory and inhibitory gradients) and Thomas' (e.g., Thomas, 1993; Thomas, Mood, Morrison, & Wiertelak, 1991) explanation (in terms of adaptation level theory; Helson, 1964). Spence (1937) assumed that an excitatory gradient forms around the S+ and an inhibitory gradient forms around the S- during simple discrimination intervention. The algebraic summation of these two gradients is considered to displace the peak of response away from the S+ in the direction opposite to the S- following intra-dimensional simple discrimination intervention. Although the S+ is located at the peak of the excitatory gradient, this peak overlaps with part of the inhibitory gradient. As a result, a location displaced from S+ that still has part of the excitatory gradient; but less of the inhibitory gradient, ends up with a higher total excitatory value (see Figure 6 for an example). In Figure 6, the S+ was 550 nm and the S- was 570 nm. The response strength on 550 nm was a summation of its performance on the excitatory gradient (a level of 90) and its performance on the inhibitory gradient (a level of -30; [minus indicating the direction]), resulting in an overall level of 60. However, the peak of the gradient was located at 540 nm, on which there was a lower level of response strength on the excitatory gradient (a level of 80) and a lower level of response strength on the inhibitory gradient (a level of -10), resulting in a level of 70. To a certain degree, Spence's theory adopted an absolute approach to discrimination learning because it assumes that organisms learn response tendencies to specific stimulus values (Cheng et al., 1997).

Thomas (1993) and Thomas et al. (1991) used Helson's (1964) adaptation level theory to explain peak shift. They suggested that the average value of stimuli presented during intra-dimensional simple discrimination intervention forms an adaptation level, or frame of reference, and that the S+ is referenced as being X units from this adaptation level. That is, the intervention recipient learns to respond to a value that is X units above or below that adaptation level. During generalization testing without feedback, the adaptation level becomes the average of the range of values used in testing. If this range differs from that used in intervention, then the code for S+ corresponds to a different stimulus value from the S+ value during intervention (Cheng et al., 1997). The rule of responding X units away from the adaptation level will result in a peak-shift effect. In this sense, Thomas' theory adopted a relative approach to discrimination learning assuming that organisms learn about relationships between stimulus values (Bizo & McMahon, 2007). Consequently, Thomas's theory emphasized the importance of the range of stimulus values used during generalization testing for the peak-shift effect.

In general, the existing evidence favors Spence's theory in nonhuman research but Thomas' theory is suggested to better explain findings from human research (Wisniewski et al., 2009). One plausible reason may be that in nonhuman research the simple discrimination intervention is typically so extensive that the adaptation level is more stable and immune from changing when new values are introduced during generalization testing (Thomas et al., 1991). In spite of the relatively greater resistance to change on the adaptation level in animals, several researchers have suggested that the potential range effects be considered in all studies of peak shift (Bizo & McMahon, 2007; Cheng et al., 1997).

Studies Addressing Generalization in Conditional Discrimination for Learners with IDD

In contrast to the evidence on generalization in simple discrimination via a generalization gradient, the parallel evidence regarding conditional discrimination is limited. Largely, studies addressing generalization in conditional discrimination for learners with IDD have been absent in the literature. One potential reason is that conditional discrimination tasks can be challenging for these individuals. Consequently, the existing literature on conditional discrimination has mainly focused on teaching discrimination skills, with little attention to given to generalization. Next the challenges with conditional discrimination tasks for learners with IDD are described, followed by a delineation of the evidence on generalization in conditional discrimination when generalization gradient has been applied, especially with learners having IDD.

Challenges Learners with IDD Have Encountered in Learning Conditional Discrimination

There is a substantial evidence base documenting that learners with moderate to severe IDD have difficulty learning conditional discrimination skills (Dube & Serna, 1998; Dube, Iennaco, & McIlvane, 1993; Ferrari, de Rose, & McIlvane, 2008; McIlvane, Dube, Kledaras, Iennaco, & Stoddard, 1990; Saunders & Spradlin, 1989; Serna et al., 1997; Zygmont, Lazar, Dube, & McIlvane, 1992). The challenges they have encountered in learning conditional discrimination may be due to their cognitive and learning characteristics as illustrated below.

Learners with IDD often have challenges with learning simple discriminations (Dickson, Wang, et al., 2006; Lovaas et al., 1971). This, in turn, makes learning higher-order conditional discrimination very problematic (Saunders & Spradlin, 1993). Second, learners with IDD can have challenges in flexibly adapting to a reversal of a mastered simple discrimination that is required for the mastery of a conditional discrimination (Dube & Mcilvane, 2002; House & Zeaman, 1962). That is, after learning to produce response B_1 not B_2 in the context of A_1, it may be difficult for learners with IDD to shift to making response B_2 instead of B_1 in the other context of A_2. Moreover, when a delay is involved in a conditional discrimination (e.g., delayed matching to sample), additional memory and/or attention demands may result (Reichle & Wilkinson, 2012). Memory represents an area in which learners with moderate to severe IDD often have impairments (Belmont & Butterfield, 1969; Henry & MacLean, 2002; Jarrold et al., 2002). Similarly, learners with moderate to severe IDD often have attention limitations (Cha & Merrill, 1994; Tomporowski & Tinsley, 1997). Both memory and attention limitations could contribute to the challenges in mastering a conditional discrimination involving a delay for learners with IDD. In summary, the existing literature documents difficulties in conditional discrimination that are associated with moderate to severe IDD (Bailey, 1981; Lovaas et al., 1971; Reichle & Wilkinson, 2012).

Evidence Addressing Stimulus Generalization Subsequent to the Acquisition of Conditional Discriminations

In his examination of conditional discriminations involving intraverbals, Axe (2008) allocated little attention to generalization. In reviewing articles between 1960 and 2015 by applying the search terms “conditional discrimination,” and “generalization gradient,” fewer than 30 refereed articles were identified that addressed the impact of conditional discrimination on generalization using a generalization gradient as a tool. Among the studies located only two were specific to learners with IDD.

In general, implementation of a generalization gradient subsequent to a conditional discrimination in which at least two response choices are available revealed a sigmoidal form for each response choice (Galizio, 1980; Heinemann & Chase, 1970; see Figure 1[f] for an example). As Figure 1(f) shows, a sigmoid form or curve is an “S” shape. In a two choice conditional discrimination for each response choice, there were generally high response rates, probabilities, or frequencies for all stimuli on the side of the teaching stimulus toward which the target response choice was reinforced. Generally, there were low response rates, probabilities, or frequencies for all stimuli on the side of the teaching stimulus toward which the target response choice was not reinforced but an alternative response choice was reinforced. In addition, there were intermediate response rates, probabilities, or frequencies for either response choice for the stimuli between the two teaching stimuli. When the gradients for two response choices are taken together, one gradient is usually the inverse sigmoidal curve of the other gradient.

The sigmoidal form of a generalization gradient subsequent to a two choice conditional discrimination suggests that after a conditional discrimination, at least two stimulus classes may emerge along a specific stimulus continuum. That is, conditional discrimination intervention may lead to the formation of an equivalent class for each of the two stimulus extremes (Arntzen, Braaten, Lian, & Eilifsen, 2011), wherein an equivalent class can be defined as containing a finite number of stimuli that produce similar results (Spradlin, Cotter, & Baxley, 1973; Sidman, 2000). In turn, a generalization gradient is useful in spotting the degree of generalization within a stimulus class, as well as, the point at which the separation of two stimulus classes becomes discriminable for a learner along a stimulus continuum. The continuum is characterized by a less discriminable center but increasing discriminable stimuli toward the two poles.

In the paragraphs that follow, two representative studies that explored the generalization performance after a conditional discrimination intervention via a generalization gradient were first described, followed by an elaboration of the two studies specific to learners with IDD.

Two representative studies addressing generalization after conditional discrimination intervention and using a generalization gradient as a measurement tool. Capehart & Pease (1968) and Galizio (1980) examined generalization gradient outcomes subsequent to the acquisition of a conditional discrimination. Capehart and Pease (1968) examined the applicability of *adaptation level theory* (Helson, 1964; Thomas et al., 1991) in explaining the generalization performance subsequent to conditional discrimination. As described in the section relevant to simple discrimination; theoretically, adaptation level theory suggests that the average value of stimuli presented during conditional discrimination intervention forms an adaptation level (or frame of reference). Each of the teaching stimuli is referenced as being X units from this adaptation level. That is, participants learn to respond to a value that is X units above or below that adaptation level. During generalization testing without feedback, the adaptation level becomes the average of the range of values used in testing. If this adaptation level differs from that formed in intervention, then the codes for the teaching stimuli change in accordance to the new adaptation level, with responding X units away from the new adaptation level. Based on this theory, Capehart and Pease (1968) hypothesized the formation of classes subsequent to conditional discrimination intervention would be impacted by the range of testing stimuli during a generalization testing.

In the study, four undergraduate male students were taught to respond “Left” to a 100 gram weight and “Right” to a 200 gram weight. The presentation of 100 and 200 gram weights were randomized. During conditional discrimination intervention, the feedback about the correctness of their choice was provided. All participants were informed that they would not be given any feedback about the correctness of their choice after the mastery that was defined as correct responses across 10 consecutive opportunities. As soon as the participants reached the mastery criterion, they were presented with a test series of seven weights ranging from 160 to 450 grams (i.e., 160, 200, 250, 300, 350, 400, and 450 grams). According to the adaptation level theory, in this study, the adaptation level during conditional discrimination was 150 grams (i.e., the average of the two training stimuli of 100 and 200 grams), whereas the adaptation level during generalization testing was around 248 grams (i.e., the average of all testing weights). Results from the generalization gradients of a sigmoidal form showed that the 250 gram weight elicited the most ambiguous response pattern (i.e., the closest to a 50-50 split of left and right responses) and that stimuli lighter than 250 grams elicited a predominately “Left” response pattern, whereas those heavier than 250 grams elicited a predominantly “Right” response pattern. The majority of the participants transposed their response to 200 grams from a “Right” response to a “Left” one in the testing phase. Also, the response latency was longest at the 250 grams and decreased directly as a function of increasing distance in either direction from 250 grams.

The findings corroborated the authors’ original hypothesis and also demonstrated that the stimulus in the middle of a continuum of testing stimuli may produce the most ambiguous response and may play a critical role in producing two response classes. Later, several researchers (Zoeke, Sarris, & Hofer, 1988; Thomas, Lusky, & Morrison, 1992) further described the findings from the study by Capehart and Pease (1968) as a shift in the *point of subjective indifference* (PSI; the stimulus value to which both responses were equally likely), from a point approximately midway between the teaching stimuli within a two-choice matching to sample paradigm to a point near the central value of the testing stimulus values (Thomas et al., 1992). To illustrate, in Capehart and Pease (1968), the PSI stimulus during conditional discrimination was 150 grams which shifted to 250 grams during generalization testing. Theoretically as the teaching values are further away from the center of the testing values during generalization testing, the magnitude of the PSI shift becomes more marked due to the larger degree of asymmetry of the test stimuli in comparison to the teaching stimuli (Thomas et al., 1992).

From a research perspective slightly different from Capehart and Pease (1968), Galizio (1980) was interested in examining whether peak shift that was often observed in the simple discrimination literature could also be found in a conditional discrimination task involving auditory stimuli. Five groups of undergraduate students aged 17 to 30 participated in this study with each group consisting of 14 participants. There were three experimental conditions, including a single stimulus intervention condition, a simple discrimination intervention condition, and a conditional discrimination intervention condition. The participants were told that their task was to judge whether a tone presented via a left or right headphone was correct, and press a key when it was judged as correct but not when it was judged as incorrect. They were also told that after they pressed the key in response to a tone, the experimenter would tell him or her whether the tone was correct or not.

Specifically, two groups of students received single stimulus intervention in which only a single tone was repeatedly presented. One group was taught to judge the tone of 1200 Hz as correct, and the other group was taught to judge the tone of 1400 Hz as correct. Two groups of students received simple discrimination intervention. One group was taught to judge the tone of 1200 Hz (S+) as correct, and the tone of 1400 Hz (S-) as incorrect. The other group was taught to judge the tone of 1400 Hz (S+) as correct, and the tone of 1200 Hz (S-) as incorrect. One group of students received conditional discrimination intervention. They were taught to judge the tone of 1200 Hz as correct and the tone of 1400 Hz as incorrect when the tones were presented in the left earphone. They were also taught to judge the tone of 1400 Hz as correct and 1200 Hz as incorrect when the tones were presented in the right earphone. Subsequent to mastery they were administered with a generalization test along a continuum of seven pure tones, ranging from 1000 to 1600 Hz, with 100-Hz steps between stimuli. No feedback was provided during generalization testing.

Results demonstrated that the single stimulus groups reached the mastery the quickest. This outcome was followed by the simple discrimination groups; and last was the conditional discrimination group. The gradients obtained for both single stimulus groups peaked at the corresponding teaching stimulus and then monotonically declined as the test stimuli moved further away from the teaching stimulus on either side of the teaching stimulus. The gradients obtained for both simple discrimination groups demonstrated peak shift or area shift. The gradients obtained for the conditional discrimination group were composed of two panels and both panels were in sigmoidal form. One panel displayed responses to tones presented in the left earphone and the other panel summarized responses to tones presented in the right earphone. Gradients peaked at 1000 Hz when tones were presented in the left earphone (where 1200 Hz was S+ and 1400 Hz was S-), but peaked at 1500 Hz when tones were presented in the right earphone (where 1400 Hz was S+ and 1200 Hz was S-). That is, the highest rate of the appropriate response occurred at the stimulus on the corresponding S+ side in a direction opposite to the corresponding S-. Thus, peak shift occurred for both conditional stimuli (i.e., tones presented in the left earphone, and tones presented in the right earphone), which was coined as *conditional peak shift* in this study. Additionally, the author found that magnitude of peak shift was larger following conditional discrimination than after simple discrimination.

The findings from this study revealed that the peak shift often observed in simple discrimination could also be obtained in conditional discrimination. Furthermore, consistent with the study by Capehart and Pease (1968), the gradient after a conditional discrimination demonstrated a sigmoidal form, reflecting the formation of two potential stimulus classes along the continuum of auditory stimuli. More importantly, the phenomenon of conditional peak shift implied that the performance within each potential stimulus class may still have some variation across the corresponding portion of the stimulus continuum. That is, for a potential stimulus class, the relevant stimuli did not produce the same amount of target response. Variation in response patterns may also occur between two potential stimulus classes. For example, one peak was found on one end of the continuum (i.e., 1000 Hz) whereas the other peak was not found on the other end of the continuum (i.e., 1600 Hz). Although the reasons responsible for these variations remain reasonably unexplored; the findings of conditional peak shift along with the PSI shift aforementioned suggest that the generalization outcomes can be somewhat manipulated and optimized by carefully selecting the teaching stimuli involved in a conditional discrimination. Next, two studies involving learners with IDD using a generalization gradient are described.

Two applications using a generalization gradient to examine generalization subsequent to a conditional discrimination among learners with IDD. Two studies (Lane & Curran, 1963; Risley, 1964) have adopted a generalization gradient to examine the influence of conditional discrimination on generalization for learners having IDD. Lane and Curran (1963) reported on three children between 7 and 10 years of age who were blind and experienced severe IDD. The learners were taught to respond differentially to two intensities of a pure tone. For one participant, a conditional discrimination was originally taught followed by the introduction of a procedure to teach a simple discrimination. For the other two participants, only a simple discrimination was taught. During the simple discrimination intervention, the participants were taught to press the right-hand button at the presence of 56-db tone, but not to press the right-hand button at the presence of 74-db tone. In the conditional discrimination intervention, at the presence of 56-db tone, the participant was taught to press the left-hand button but not the right-hand button; at the presence of 74-db tone, the participant was taught to press the right-hand button but not the left-hand button. Incorrect responses resulted in no effect. Each correct response was reinforced until the learner responded without error across two intervention series for the simple discrimination intervention and one intervention series for the conditional discrimination intervention. Then the generalization testing was implemented involving 11 different intensities (50, 53, 56, 59, 62, 65, 68, 71, 74, 77, and 80 db) with each of them presented 10 times. Two weeks later the generalization testing procedures were conducted again to examine the maintained generalization performance.

Results demonstrated that subsequent to a simple discrimination intervention, the generalization gradient of response frequency as a function of tone intensity was somewhat monotonically increasing across the test stimuli and demonstrated a peak shift. This is consistent with the findings for an intensity gradient (Brennan & Riccio, 1973; Pierrel & Sherman, 1960; Scavio & Gormezano, 1974). Performance in generalization gradients subsequent to the conditional discrimination intervention was plotted as the relative frequency of left- and right-hand responses evoked by 10 presentations of the corresponding tone intensity, respectively. The two gradients were in a reversed sigmoidal format and were complementary to each other, with the intersection point located between the two teaching stimuli (i.e., 56-db and 74-db tones), that is, around 71-db tone. Additionally, they were both non-monotonic change in response probability and not symmetrical (i.e., a higher proportion of response to the right-hand responses). The form of the gradient remained stable two weeks after the intervention completed. On the basis of these results, the researchers concluded that the generalization gradients obtained for learners with IDD were similar to those obtained with normal college students in their previous study (Cross & Lane, 1962). They summarized that learners with IDD formed two stimulus classes along a continuum of generalization testing stimuli after a two-sample two-choice conditional discrimination intervention.

Similarly, Risley (1964) was interested in investigating the generalization performance across a wide range of values on the auditory frequency dimension following a conditional discrimination for four male adolescents with IDD by use of a generalization gradient. The participants were taught with a two-sample two-choice matching to sample task that involved pulling the right bar in the presence of the 600 cps tone and pulling the left bar in the presence of the 1900 cps tone on a variable ratio reinforcement schedule. The intensity of the S^D tones was increasingly varied between S^D periods until the participants were responding equally and accurately to S^D tones of the two frequencies (600 and 1900 cps) which varied in intensity between 50 db and 64 db. Sessions were approximately 20 min at the start of intervention and gradually increased to approximately 60 min by the end of the study, and occurred two to four times per week for up to eight months. Subsequent to mastery, the participants received generalization testing, in which the tones ranged from 300 to 6000 cps in roughly logarithmic steps as follows: 300, 400, 500, 600, 900, 1200, 1500, 1900, 2600, 3500, 4600, 6000 (which includes the two S^D tones of 600 and 1900 cps; 12 stimuli in total). After four sessions of testing, the session length was doubled and the range was extended from 160 to 12,000 cps (16 stimuli in total).

Performance in the generalization gradient was graphed as the proportion of total responses on the right bar, which was the exact complement of the proportion of responses on the left bar (not depicted in this study). The 600 cps frequency was the S+ and the 1900 cps was the S- for responding on the right bar. All participants responded entirely on the right bar during the probes of the 600 cps tone and entirely on the left bar during the probes of the 1900 cps tone. Two participants also responded entirely on the right bar on all frequencies below 600 cps and entirely on the left bar on all frequencies above 1900 cps. The other two participants showed more variability, but responded predominantly on the right bar on the lower frequencies and on the left bar on the higher frequencies. There was a systematic shift in response distribution between the two S^D points and three participants showed a reversal of their trend at 1500 cps. In addition, the latency gradient was bimodal in format in general, with higher latency at the two S^Ds.

Overall the two studies for learners with IDD focused only on the auditory stimuli with respect to the sound intensity or tone frequency dimensions. In line with the study by Lane and Curran (1963), the results from Risley (1964) are consistent with the proposition that training stimuli involved in the conditional discrimination intervention enhanced the separation of the generalization testing stimuli into two major equivalent classes for the two response choices (Arntzen et al., 2011).

In spite of the general findings on the generalization gradient performance obtained after a conditional discrimination intervention, rarely has research examined whether and how specific intervention parameters influence the outcome of conditional discrimination task with respect to generalization. The specific outcomes of interest may include, but not limited to; maximizing the desired generalization and minimizing the undesired generalization within each equivalent class, and optimizing the separation of different equivalent classes.

Among a group of intervention parameters involved in a conditional discrimination (e.g., the selection of teaching stimuli, intervention intensity, reinforcement schedule, the range of test stimuli in generalization testing); the specific parameter of the "difference between teaching stimuli" in a conditional discrimination is the focus the current study. Compared to the well-established evidence surrounding the impact of the difference between teaching stimuli in simple discrimination on the generalized outcomes (see Ghirlanda & Enquist, 2003, for a review), the parallel evidence for conditional discrimination remains largely sparse, let alone for learners with IDD. To date, only three studies on this regard were found and would be delineated in the section to follow.

Studies Examining Influences that Incremental Differences Between Teaching Stimuli in Conditional Discrimination Intervention Have on Generalization

Three studies (Eckerman, 1970; Horner et al., 1986; LaBerge, 1961) have examined the effects of differences between teaching stimuli on generalization performance after teaching a conditional discrimination. LaBerge's (1961) study was the first attempt to using a generalization gradient to explore how variations between teaching stimuli in a conditional discrimination may impact the generalized outcomes. In this study, undergraduates were taught to press lever A_1 when a rectangle of white light positioned at S_1 spot along a vertical axis occurred, and press lever A_2 when a rectangle of white light positioned at S_2 spot along the same vertical axis occurred. Sixteen participants were assigned to each of the three conditional discrimination conditions with different intervals between teaching stimuli, that is, 2, 4 and 32 inches between S_1 and S_2, respectively. The intervals between teaching stimuli all had the same midpoint along the vertical axis. Participants who met the mastery criterion (i.e., at least 13 correct responses in the 7th intervention block) subsequently received generalization tests on seven stimuli lying between the teaching stimuli, S_1 and S_2, along with S_1 and S_2. Each intervention block consisted of 17 teaching opportunities with the presentation of teaching stimuli randomized. Each generalization testing block consisted of five opportunities for S_1, five opportunities for S_2, and one opportunity for each of the seven test stimuli, resulting in a total of 17 opportunities per block. There were 35 generalization testing blocks for each participant, totaling 595 testing opportunities. In addition, each participant were instructed to press a lever when a white light appeared and were told that a red light above one of the levers would flash to indicate which lever he or she should have pressed on that opportunity. They were also told that when no red light followed a response, such opportunities were testing opportunities wherein the experimenter wanted to know which lever should be pressed for each opportunity.

Results revealed that all participants in the 32-inch condition met the mastery criterion, whereas five participants in the 4-inch condition and seven participants in the 2-inch condition did not meet the criterion. Additionally, the generalization gradient for each group of teaching condition was depicted as the proportion of A_1 responses across all test stimuli (i.e., seven novel stimuli and the two original teaching stimuli S_1 and S_2). The group curves were of sigmoid shape, with plateaus of varying length near the teaching stimuli. Furthermore, the magnitude of the interval between teaching stimuli had a highly significant effect upon the slopes of the gradients. That is, as the interval decreased, the slope became larger, suggesting a higher degree of differentiation in terms of the responding to the two levers, and a larger degree of generalization of the responding to one of the teaching samples (e.g., S_1) to other novel test samples close to the side of the corresponding teaching sample (i.e., S_1 in this case). Herein, the slopes of the gradients were obtained by taking the last stimulus point whose response proportion was less than 10% and the first stimulus point whose response proportion was greater than 90% and then computing the slope of the line connected by these two stimulus points. Besides, group latency gradients were also plotted indicating a monotonic increase in latency as the test stimulus was further away from the teaching stimulus across all three conditions.

The findings from this investigation suggested that the difference between teaching stimuli in a conditional discrimination intervention likely be a functional variable that could exert influence on the generalized outcomes in addition to the mastery of discrimination. Moreover, the study operationalized an approach of calculating the slope of a generalization gradient subsequent to a conditional discrimination intervention to quantitatively reflect the degree of the separation of two target responses across a continuum of test stimuli. Nevertheless, in this study the test stimuli covered in three conditional discrimination conditions were not exactly the same, which may be a threat to its internal validity. Also, the maintenance of generalization performance was not explored.

Since LaBerger's (1961) investigation, little research has been implemented to further examine the effects of the difference of teaching stimuli in conditional discrimination on generalization performance using a generalization gradient. One exception was the study by Eckerman (1970) who explored this area by examining how a variation in the requirement for a sample-specific observing response, which referred to making an overt response toward each sample prior to making choices for the specific sample (Doughty & Hopkins, 2011), influenced the generalization outcomes in terms of the choice responses. Three groups of pigeons with five pigeons in each group were taught a conditional discrimination in which the samples were different colored wavelengths and the choices were two lines (vertical and horizontal). For each of the three groups, during each teaching opportunity, one of two sample colors (506 nm or 583 nm) was presented, and a sample-specific observing response was required (i.e., pecking in place I of a long slot for sample color 506 nm; pecking in place II of the slot for sample color 583 nm) before choices (vertical and horizontal lines) were presented. In particular, the difference between place I and place II required for the observing responses varied among these three groups of pigeons, by 0, 3, and 6 inches for Groups 1, 2, and 3, respectively. Once the choices were presented, observing responses no longer had any effect. At the presence of the 506-nm sample, a response to the horizontal line was correct, while at the presence of the 583-nm sample, a response to the vertical line was correct. The first peck at one of the choices terminated the opportunity and produced either reinforcement or blackout. After the birds reached a high level of accuracy in the conditional discrimination task, generalization testing was conducted in extinction using both novel samples and novel choices. Any of five color stimuli (506, 525, 548, 565, and 583 nm) was presented as the sample, and line orientations of 0, 30, 45, 60, and 90 degrees were presented in pairs as the choices.

Results demonstrated that the group of pigeons that were required to make the sample-specific observing response involving the largest difference (i.e., the widest separation between place I and place II) reached the mastery criterion most rapidly. Additionally, their performance yielded the highest level of accuracy. Generalization gradients for both the observing response and the choice response were obtained. The gradient for the observing response was depicted as the median response location along the slot across the five test stimuli. For Group 1, the median response location was fairly constant across all test stimuli. For Group 2, the median response location changed gradually across the test stimuli, between the corresponding place I and place II on the slot. For Group 3, the median response location was generally within either one place or the other. For Groups 2 and 3, there was a shift in response locus which took place between 525 nm and 548 nm for most birds, which was more marked in Group 3. The generalization gradient for the choice response for each of the three groups was graphed as the probability of response to each choice (i.e., each line orientation) for each of the five test stimuli. Consistent with the findings for the observing response, across the test stimuli the probability of a response to each line orientation was generally less differentiated in Group 1 than those in Groups 2 and 3. Moreover, for Groups 2 and 3 the shift in response probability also occurred between 525 nm and 548 nm, reflected by the point of change in sign of the slope along the gradient. To be more specific, at the test stimuli of 506 nm and 525 nm, an overall higher probability of responding was found for line orientations closer to the end of 0 degree. At the test stimuli of 565 nm and 583 nm, an overall higher probability of responding was found for line orientations closer to the end of 90 degree.

Although the difference between the teaching stimuli in this study was manipulated by mediating the difference involved in the sample-specific observing responses, the findings revealed that as the difference increased, not only the discrimination learning became easier for the pigeons, but both the observing response and the choice response became more differentiated between that appropriate for one of the teaching stimuli and that appropriate for the other. In other words, as the difference increased, there was a larger degree of generalization of the instructional control exerted by one of the teaching samples to other novel test samples close to the side of the corresponding teaching sample. However, it should be noted that the findings from this study were more clearly exhibited when comparing the 0-inch condition and the 3- or 6-inch conditions, but not quite clearly exhibited when comparing the 3-inch and 6-inch conditions. Also, the findings from this study were not in concordance with those from LaBerge (1961), which may be owing to the procedural differences used in these two studies, and suggested the need for more research in this field.

Unlike the two previous studies, Horner et al. (1986) were interested in how the difference between teaching stimuli impacted learners' generalization of a rejecting skill. During grocery shopping they taught five adults with IDD to select each of 10 items that matched a corresponding picture card. Additionally the participants learned to reject a set of items that did not match picture cards representing the objects that included; (a) items that were maximally different from those to be selected (i.e., maximal-difference intervention condition), or (b) the items that were minimally different from those to be selected (i.e., minimal-difference intervention condition). Within a split-multiple baseline design across participants, three participants received the maximal-difference intervention condition, followed by the minimal-difference intervention condition, and the other two participants only received the minimal-difference intervention to control for a potential multiple intervention interaction. Intervention sessions to criterion were reported to be equivalent and both intervention conditions were sufficient for the learners to correctly select the 10 items taught in a novel store. The investigators reported that the minimal-difference condition was necessary for the learners to correctly reject untrained items in the novel store. Since the differences between teaching items were not clearly specified, it was not possible to quantify the differences along certain stimulus dimension(s) via a generalization gradient to gain a complete picture of learners' performance. This is particularly the case with respect to a concurrent picture of both potential over- and under-generalization across a continuum of stimuli. It was also not possible to examine the generalization effects over time since no maintained generalization performance was reported in this study. Similarly, it was not possible to examine how small the difference between teaching stimuli during conditional discrimination intervention was necessary for learners with moderate to severe IDD to minimize under-generalization of the newly taught rejecting skill.

Summary

The purpose of this chapter was to provide a review of the empirical evidence surrounding the issue of generalization in conditional discrimination for learners with IDD. A particular focus was placed on the parameter of difference between teaching stimuli in conditional discrimination, which is necessarily brief because there are so few studies in which both the parameter of difference between teaching stimuli and the tool of a generalization gradient are taken into consideration. Taken together, there does not appear to be solid evidence addressing the effects of parametric alterations in conditional discrimination on generalized outcomes for learners with IDD. Consequently, additional research in this area is warranted.
In the next chapter, the methods of the experiment will be presented. The experiment examined how differences between teaching stimuli in a conditional discrimination task impacts the generalization performance among five learners with moderate to severe IDD, with using the tool of a generalization gradient. Prior to the description of the experiment, a pilot study will be first briefly described, which examined the performance of typically developing adults to validate the color stimuli that were selected to be used in the experiment.

CHAPTER III: METHOD

In this section, an overview and outcomes of a pilot study conducted to inform methodological components of the primary experiment that was the focus of this study are described. A complete description of the pilot study appears in Appendix of this dissertation.

Pilot Study: Experimental Stimuli and Exploration of Response Mode for Dependent Measure

Purpose

This pilot was designed to; (a) validate the color stimuli to be used in the current experiment, and (b) explore the use of two different response modes used to obtain the dependent measure, and (c) obtain the generalization gradients averaged across typically developing adults along each of three final color sets for two response modes. Color stimuli were utilized because they have successfully been used in the stimulus generalization literature in single stimulus learning (Guttman & Kalish, 1956), simple discrimination learning (Hanson, 1959), and conditional discrimination learning (Livesey & McLaren, 2009; Thomas & Curran, 1988).

With respect to the validation of color stimuli, the pilot study was aimed at identifying the color stimulus among each of three color sets that was *point of subjective indifference* (PSI; i.e., performance closet to 50% chance level in terms of judging as a specific color; Thomas et al., 1992). Subsequently, with the PSI stimulus determined, the same number of color stimuli on both sides of the PSI stimulus for each of the three color sets was identified, which resulted in the final three color sets to be used in the experiment to be reported.

The second purpose of the pilot was to examine whether two different response modes would produce similar results. One response mode required the participants to make a forced choice between two color categories (e.g., "Blue" or "Green"), whereas the other response mode required the participants to make a forced choice between "Yes" and "No" in response to a question like “Is it green?” Hardin and Maffi (1997) suggested that a forced choice yes/no response eliminates the need for a linguistic response. However, others have documented that a yes/no response can be cognitively more demanding for learners with moderate to severe IDD (Cederborg, La Rooy, & Lamb, 2008). To determine whether both response modes would yield similar results both response modes were taken into account in this pilot investigation.

Finally, the generalization gradients averaged across typically developing adults along each of three final color sets for two response modes were determined. The resulting “typical” generalization gradients were compared with the ideal generalization gradient by use of the indicator of the residual sum of squares (RSS). To calculate the RSS value for each color set generalization gradient in each response mode condition, the predicted value from the ideal generalization gradient was first subtracted from the observed value from the participant generated generalization gradient. Subsequently, this result was squared. Finally the sum of all the squared outcomes represented the RSS value. For each color set, the RSS values for generalization gradients under two response modes were further averaged to indicate the overall discrepancy between the typically developing adults’ performance and the ideal performance with both response modes taken into account. The resulting RSS values were used to gauge whether there was room for improvement and instruction based on the baseline performance of learners with IDD in the primary experiment.

Main Outcomes and Implications for the Primary Experiment

The methods and results of the pilot investigation are reported in detail in Appendix. In general results suggested that the participants showed most uncertainty between two choices for each of two response modes on the color stimulus of 499 nm on the blue-to-green color set, the color stimulus of 594 nm on the yellow-to-orange color set, and the color stimulus of 631 nm on the orange-to-red color set. These three color stimuli (i.e., 499, 594, and 631 nm) were designated as the final PSI stimuli for each of the three color sets. With the PSI designated for each color set, results suggested that the an average interval of 3.44 nm on the wavelength and an average interval of .03 on the luminosity[1] value between two adjacent color stimuli for each color set seemed to be practically feasible in making the adjacent stimuli sufficiently discriminable for the participants to make judgments. Furthermore, Table 8 shows the final three color sets used in the primary experiment, with each final color set consisting of nine color stimuli (i.e., the PSI stimulus and four stimuli on each side of the PSI stimulus).

Secondly, the findings from the pilot study suggested that in spite of some differences, participants' performance for the two response modes on each of the three color sets was very consistent. It was noted that response latency was greater for forced choice between "Yes" and "No" on two PSI stimuli identified (i.e., 594 nm for the yellow-to-orange color set, and 631 nm for the orange-to-red color set). Additionally, on the yellow-to-orange color set participants' performance between the two adjacent stimuli of 594 nm and 597 nm was more clearly differentiated when the forced choice yes/no response mode was required. Both of the two response modes were judged to be empirically reasonable candidates for the adult participants to make judgment on color categories. Although the performance on the yes/no response mode

---

[1] Luminosity is the measure of the effectiveness of lights of different wavelengths (Sharpe, Stockman, Jagla, & Jägle, 2005). The term was introduced by the International Lighting Commission (Commission Internationale de l'Eclairage or CIE) to reflect the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). It is based on subjective judgements of which of a pair of different-colored lights is brighter, to describe relative sensitivity to light of different wavelengths. The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_{\lambda}}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_{λ} is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005). The value is close to 1 when the wavelength of the light is 555.016 nm, to which humans are most sensitive. The value decreases when the wavelength of the light is higher or lower than 555 nm. Notably, the visible wavelengths for human beings range from about 390 to 700 nm. The original sensitivity values for each wavelength was found on http://web.archive.org/web/20070314050445/http://www.cvrl.org/database/data/lum/ssvl2e_1.txt

seemed to be more differentiated, the yes/no response mode can be cognitively more demanding for learners with moderate to severe IDD and the results may be less accurate (Cederborg et al., 2008). Consequently, a combination of the two response modes was developed for learners with moderate to severe IDD in the experiment (to be further described below in the methods for the current experiment).

Finally, in this pilot, the typically developing adults' performance was not exactly the ideal performance (see Figure 8 described in the Appendix). As this figure shows, in the ideal generalization gradient, the percent of independent selection of a target choice was 100% in the presence of the PSI stimulus and each of the four stimuli that belonged to the designated stimulus Class A. Meanwhile, the percent of independent selection of the same target choice was 0% in the presence of each of the four stimuli that belonged to the designated stimulus Class B. Herein, for the stimuli within the designated stimulus Class A, the correct response choice is the same as the one to which the PSI stimulus corresponded (i.e., the target choice). For the stimuli within the designated stimulus Class B, the correct response choice should be the other option. More specifically, in this study on a specific color set, one specific color choice was designated as the correct response given the PSI stimulus and the stimuli within the stimulus Class A; whereas the other color choice was designated as the correct response given the stimuli within the stimulus Class B. The RSS values obtained reflect the reasonable room of deviation from the ideal performance. The lower the RSS is, the closer the actual performance is to the ideal performance. As Table 8 shows, the RSS value of the generalization gradients averaged across two response modes was 0.11 on the blue-to-green color set, 0.64 on the yellow-to-orange color set, and 0.42 on the orange-to-red color set. With color stimuli chosen along with response mode, methods for implementation of the primary study will be subsequently described.

Acquisition and Generalization of a Conditional Discrimination Among Participants with Moderate to Severe IDD

This experiment investigated whether and how the degree of discrimination and generalization of color stimuli would vary as a function of the difference between the teaching stimuli in a conditional discrimination (i.e., a maximal difference, and a minimal difference) among learners with moderate to severe IDD. An adapted alternating treatment single case experimental design (Kazdin, 2010) was employed involving three experimental conditions.

The three distinct experimental conditions consisted of two conditional discrimination conditions, that differed only in the degree of color difference (maximal or minimal) between the two teaching stimuli (i.e., S_A and S_B), and a simple discrimination condition. The simple discrimination condition, as frequently used in basic discrimination research to examine whether similar findings obtained on a simple discrimination condition were able to be found on a conditional discrimination condition (e.g., Blough, 1973; Galizio, 1980; Heinemann & Chase, 1970; Thomas et al., 1992; Thomas, McKelvie, Ranney, & Moye, 1981), was included in the current study as an experimental condition.

In all three conditions, the first teaching stimulus (S_A) was the PSI stimulus for the respective color set. In the ***simple discrimination condition***, the PSI stimulus served as the only teaching stimulus (S_A) and the designations of the two choices as S+ and S- (which were the two anchor colors; e.g., blue and green) remained the same across opportunities.

In the ***maximal-difference conditional discrimination condition*** the S_A was the PSI stimulus of a specific color set, and the S_B was the stimulus most different from the S_A on the color set. In the ***minimal-difference conditional discrimination condition*** the S_A was also the PSI stimulus of a color set, but the S_B was the stimulus in the middle of the PSI stimulus and the stimulus most different from the S_A on the color set.

The three different color sets were finalized through the pilot study. As Table 8 shows, the wavelengths of the color stimuli in the blue-to-green color set were 483, 489, 493, 496, 499, 502, 504, 506, and 508 nm; and the corresponding luminosity values were .2002, .2361, .2699, .3027, .3411, .3854, .4182, .4533, .4904. The wavelengths of the color stimuli in the yellow-to-orange color set were 580, 584, 588, 591, 594, 597, 600, 603, and 606 nm; and the corresponding luminosity values were .8822, .8496, .8130, .7804, .7450, .7074, .6688, .6299, and .5899. The wavelengths of the orange-to-red color set were 619, 622, 625, 628, 631, 634, 637, 641, and 645 nm; and the corresponding luminosity values were .4151, .3784, .3426, .3061, .2709, .2386, .2094, .1754, and .1458.

The assignment of three color sets to the three intervention conditions was counterbalanced across participants in the experiment. Table 9 presents a schematic diagram to show an example of the designation of teaching stimuli across the three experimental conditions. The tabled example assumes the blue-to-green color set was assigned to the simple discrimination condition, the yellow-to-orange color set to the maximal-difference conditional discrimination condition, and the orange-to-red color set to the minimal-difference conditional discrimination condition. In particular, the four stimuli on either side of the PSI stimulus were assumed to belong to two stimulus classes. For the stimuli within stimulus Class A, the correct response choice was the same as the one in the presence of the PSI stimulus. For the stimuli within stimulus Class B, the correct response choice was the other option.

Participants

Five children with Down syndrome and moderate to severe intellectual delays participated. The experimenter obtained consent from each participant's parents. All participants were recruited through a regional Down syndrome parent advocacy group. The participants met the following inclusion criteria.

Inclusion criteria. The inclusion criteria included: (a) a diagnosis of moderate to severe disability, evidenced by performance at least two standard deviations below the mean on at least one subscale of the *Mullen Scales of Early Learning, AGS Edition* (MSEL-AGS; Mullen, 1995), (b) normal color vision by passing the color vision screening using the *Color Vision Testing Made Easy®* test (CVTMET; Waggoner, 1994), (c) at least one year delay in receptive vocabulary skills, evidenced by the performance on the *Receptive One-Word Picture Vocabulary Test, Third Edition* (ROWPVT-3; Brownell, 2000), and at least one year delay in expressive vocabulary skills, evidenced by the performance on the *Expressive One-Word Picture Vocabulary Test, Third Edition* (EOWPVT-3; Brownell, 2000), (d) demonstrated ability to follow one-step directives such as “Come and sit down,” “Give it to me,” and “Push the button,” (e) inability to discriminate between target colors demonstrated by baseline data, and (f) no problem behavior that would interfere the instruction as judged by the experimenter. The paragraphs to follow provided a description of the screening assessments involved in these inclusion criteria.

Developmental assessment. The Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995) was administered to assess the developmental functioning of each participant. The Mullen scales encompass five subscales: Gross Motor, Visual Reception, Fine Motor, Receptive Language, and Expressive Language. A doctoral-level research associate conducted the assessment for each participant. It took about 45 to 60 min for each participant to complete the assessment.

Adaptive functioning assessment. The Parent/Caregiver Rating Form of Vineland Adaptive Behavior Scales, Second Edition (Sparrow, Cicchetti, & Balla, 2005) was administered to assess the adaptive functioning of each participant. It includes five subscales: Communication, Daily Living Skills, Socialization, Motor Skills, and Maladaptive Behavior. One of the parents of each participant completed the rating form in about 30 min.

Communication skills assessment. The Receptive One-Word Picture Vocabulary Test (ROWPVT; Brownell, 2000) and the Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000) were administered to assess the receptive and expressive communication skills for each participant. Both assessments are individually administered, norm-referenced assessments applicable to people aged from 2 years old to over 80 years old. One experimenter whose native language was English conducted the assessments for each participant and each assessment took about 20 min for each participant.

Color vision screening. The Color Vision Testing Made Easy® test (CVTMET; Waggoner, 1994) was implemented to examine each participant's detection of colors. The CVTMET allows; (a) the use of shapes and figures familiar to young children, thereby eliminating the need for a verbal response, (b) adaptability to a tracing, matching, or forced-choice format, and (c) easy administration. It consists of one demonstration card and nine test cards displaying a circle, star, and/or square. All cards contain two shapes except for the eighth card, which has three shapes. The passing criteria are: (a) correctly identifying eight out of nine circles (there is one circle on each of the cards), and (b) not making any mistakes on cards 3, 4, and 5.

To demonstrate that the participant understood the task instructions and requirements a demonstration was provided. The experimenter presented the demonstration card to the participant and asked "Can you find a big circle (or ball) on this card (or picture)?" If the participant pointed to or traced it within 3 s, the experimenter provided the specific feedback and praise, "Yes, that's the circle (or ball). Great work." If the participant did not produce a response within 3 s, the experimenter provided a further directive, like "use your finger to tell me (or us) where the circle (or ball) is." If the participant pointed to or traced it within 3 s, the experimenter provided the feedback and praise, "Yes, that's the circle (or ball). Great work." If the participant still did not produce the correct response, the experimenter provided hand-over-hand guidance to help the child trace out the circle (or ball), while saying “Yes, here is the circle (or ball).”

Each of nine test cards was presented individually to each participant. The discriminative stimulus for each opportunity was, “Can you find the circle (or ball) on this card (or picture)?” If the participant emitted a response within 3 s, it was recorded as correct or incorrect. If the participant did not make a response within 3 s, the experimenter provided the further directive, “use your finger to tell me where the circle (or ball) is.” If the participant still did not make a response within 3 s, the response was recorded as incorrect. If the participant identified eight of nine cards, he or she was qualified to proceed in the investigation. The profile of each participant was described in the following.

Participant profiles. Each of the five participants was diagnosed with Down syndrome. Participants had a mean chronological age of 4 years and 5 months (range: 4 years 1 month to 4 years 9 months) at the beginning of participation in the experiment. All attended schools in their respective home school district, and received occupational therapy, physical therapy, and speech and language services in their schools and/or in private settings. All participants were ambulatory and passed the color vision screening (as described above). They were all at least three standard deviations below the mean on at least one subscale of Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), which resulted in the inability to calculate a composite T score and corresponding derived standard scores. This result placed their performance into the range commensurate with a moderate to severe delay.

Table 10 shows the assessment results for all five participants. On the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000), their receptive vocabulary skills were all below the 10th percentile, with an average age equivalent of two years three months. On the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-

3; Brownell, 2000), their expressive vocabulary skills ranged from being non-testable to performing at the 4th percentile, with an average age equivalent of two years and one month.

Bruce. Bruce was a four year nine month old Caucasian-Asian boy who was diagnosed with Down syndrome. Approximately two months before his participation in the study, Bruce began attending a community preschool classroom that met three afternoons a week. Prior to that, he received early childhood special education services in a mainstream classroom that met three mornings a week. On the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000) Bruce's standard score was lower than 55, with an age equivalent of one year and five months. His standard score on the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000) was lower than 55, with an age equivalent of one year and 10 months. Both his receptive and expressive vocabulary skills were below the 1st percentile rank. On the Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), he had a standard T score lower than 20 on all four subscales (i.e., Visual Reception, Fine Motor, Receptive Language, and Expressive Language), with an age equivalent of two years and seven months. On the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al., 2005) he had a standard score of 64, which placed him at the 1st percentile rank.

Bruce's Individualized Education Plan (IEP) completed at four years and 11 months of age, reported that he used a combination of sign and spoken language at home and sometimes used the two interchangeably when speaking. His articulation and intelligibility were impaired which resulted in significant intelligibility challenges, particularly with unfamiliar individuals. His parents reported that he had begun to use three word utterances to request for food. Per his teacher's report, Bruce had minimal spontaneous speech with peers, but enjoyed playing alongside them. When he was asked a question, he often repeated the question rather than answering. He was reported to be able to follow one-step directives and sometimes he was able to follow two-step directives.

Gabi. Gabi was a four year one month old Caucasian girl with Down syndrome. She attended a 3-4 year old preschool program two days a week. Gabi produced a standard score of 73 on the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000) which yielded an age equivalent of two years and five months and placed her at the 4th percentile rank. Her standard score on the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000) was 68, which yielded an age equivalent of two years and three months and placed her at the 2nd percentile rank. On the Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), she had a standard T score lower than 24 on all four subscales (i.e., 20 on the Visual Reception subscale, 20 on the Fine Motor subscale, 23 on the Receptive Language subscale, and lower than 20 on the Expressive Language subscale), with an age equivalent of two years and six months. On the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al., 2005) she had a standard score of 94, which placed her at the adequate level of 34st percentile rank.

Her IEP completed at the time of the investigation reported that Gabi was occasionally able to produce spontaneous two-word combinations. She had a productive vocabulary of 30 different spoken words based on her mom's report. Overall, Gabi used spoken words more than gestures or signs when spontaneously communicating with other people such as her speech therapist. Gabi's articulation and intelligibility were impaired making it difficult to understand her unless the context was known. With respect to comprehension, she was able to follow one-step directives. Gabi was not yet able to answer WH questions independently. She participated in parallel play and occasionally played with peers.

Bryan. Bryan was a four-year one-month old Caucasian boy with Down syndrome. He attended an early child special education classroom five mornings per week. Bryan's standard score on the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000) was 76, with an age equivalent of two years and seven months and a 5th percentile rank. His standard score on the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000) was lower than 55, which yielded an age equivalent of one year and three months and placed him lower than 1st percentile rank. On the Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), he had a standard score no higher than 30 on all four subscales (i.e., 20 on the Visual Reception subscale, lower than 20 on the Fine Motor subscale, 30 on the Receptive Language subscale, and lower than 20 on the Expressive Language subscale), with an age equivalent of two years and five months. On the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al., 2005) he had a standard score of 62, which placed him at the 1st percentile rank.

His IEP completed at his age of three years 11 months reported that Bryan had difficulty with his overall expressive language and speech-articulation skills. He used some spoken words, signs, pictures, and gestures to communicate with others. He was able to produce /m/ and /b/ sounds in the initial position of CVC and CVCV approximations consistently when presented with pictures. However, his intelligibility was limited as a result of his impaired articulation. Bryan was not yet able to produce two-word utterances consistently. He was able to respond to WHAT and WHO questions consistently with spoken language. He was able to follow common one-step directive but had difficulty with following novel directives. He was reported to engage in parallel play, and with the prompt from the adults, he was able to initiate and enter play with peers by handing objects and/or verbalizing “play?” and play for about three to five min.

James. James was a four year five month old Caucasian boy with Down syndrome. He attended an early childhood special education classroom four mornings per week. James’s standard score on the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000) was lower than 55, which produced an age equivalent of one year and six months and placed him lower than 1st percentile rank. He was not able complete the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000) which was the result of very limited spoken word capability. On the Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), he had a standard score lower than 20 on all four subscales, with an age equivalent of one year and nine months. On the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al., 2005) he had a standard score of 65, which placed him at the 1st percentile rank.

His IEP completed at three years and eight months of age, reported that he only verbally produced approximations of several words including water, duck, up, bye, quack-quack and some other animal sounds. He did not use many different consonants or consonant-vowel combinations. He understood and spontaneously used approximately 30 sign approximations to express his needs and thoughts, including more, all done, thank you, hi, ball, eat, food items such as cookie, cracker, milk, water, animals such as horse and cow. Approximately half year prior to participating in the study, he also started using a mini-iPad with a communication app on it. On the app, each page consisted of no more than 12 picture symbols. Little further information was available at the time of the investigation regarding speech generating device use other than he needed other people to navigate between pages on the app. He was able to follow simple one-step directives but had difficulty with following novel directives.

Corbin. Corbin was a four-year eight-month Caucasian-Asian boy with Down syndrome. He attended an early childhood special education classroom twice per week and a creative play class twice a week. Corbin obtained a standard score of 79 on the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3; Brownell, 2000), with an age equivalent of three years and three months and an 8th percentile rank. His standard score on the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000) was 73, with an age equivalent of three years and a 4th percentile rank. On the Mullen Scales of Early Learning, AGS Edition (MSEL-AGS; Mullen, 1995), he had standard T scores of 30 on the Visual Reception subscale, lower than 20 on the Fine Motor subscale, 23 on the Receptive Language subscale, and

20 on the Expressive Language subscale. His age equivalent score on the Mullen was three years. On the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al., 2005) he obtained a standard score of 73, which placed him at the 4th percentile rank.

Corbin used spoken words, signs, and gestures to express his wants and thoughts. He was able to understand and express about 150 signs based on his grandma's report. According to his IEP completed at his age of four years and eleven months, he was able to answer social questions such as "What's your name?", "How old are you?", and "Are you a boy or girl?" appropriately. As his IEP indicated, Corbin inconsistently responded to yes and no questions. However, according to his IEP, he was able to answer a variety of "what", "where", and "who" questions when a visual picture was present to provide context for the question. When commenting about activities or things he saw in the environment, he occasionally spontaneously combined words but often used one-word phrases. He was able to ask for help with a verbal prompt from an adult. Corbin's articulation and intelligibility were relatively intact and most of his speech can be understood by both familiar and unfamiliar people especially when the context was known. He followed one-step directive and some familiar two-step directives.

Setting

Each participant's home served as the setting for the experiment. A table in each child's living or dining room was used to display experimental materials along with a laptop computer. The experimenter(s) sat or kneeled beside the participant during assessment and intervention activities. Sessions were held twice per week for James (approximately an interval of three days between two sessions) and once per week for the other four participants, at a time of the day chosen by participants' parents. Each session lasted about half an hour to 1.5 hr, with an approximate average of one hour per session. All five participants finished the experiment within 12 weeks ranging from 8 to 12 weeks, within two to four months with some holiday breaks taken into consideration.

Materials

Laptop-computer. A 17-inch Toshiba® laptop was used to present all stimuli on a white background through the Presentation® software (Neurobehavioral Systems, 2004). The screen brightness was set as 50%.

Switches. Two identical black push-plate switches Talking Icon® (Enabling Devices, 2011) were used to allow the participants to make selections between two picture choices throughout the study. The two switches were connected to the computer by a USB-based converter. Each switch had a 2.25 inch × 2.25 inch square surface with a 45° tilt. Two picture choices that varied across the phases of the study were printed on glossy paper approximately 2.2 inches by 2.2 inches, and were inserted into the surface of the two switches respectively.

Stimuli used in the pretraining tasks. The identity matching to sample pretraining tasks utilizing familiar pictures were designed to ensure the participants understood how to complete a basic two-sample two-choice matching to sample task. Four pairs of pictures representing objects or cartoon characters obtained from Google© search engine (i.e., bird™ versus pig™; apple™ versus pumpkin™; Mickey Mouse™ versus Donald Duck™; and Thomas Train™ versus Sponge Bob™) were used as samples presented on the laptop. These pictures were arbitrarily selected and confirmed with the participants' parents to make sure their children were familiar with these objects or cartoon characters. Each of the four pairs of object or cartoon characters was printed on the glossy paper in the size of 2.2 inches by 2.2 inches. The resulting four pairs of paper-version objects or cartoon characters were inserted into the screens of the two switches respectively as the choices for the participants to select.

The color circles that would be placed on the two switches respectively as the choices for the participants to select in the experiment were referred to as the "perfect" color circles. The identity matching to sample pretraining tasks utilizing "perfect" color circles were designed to ensure that the participants were able to discriminate two “perfect” colors for each color set. In these tasks, the “perfect” color circles as samples presented on the laptop were derived from the validated color sets through the pilot investigation. As Table 11 shows, for the blue-to-green color set, the “perfect” blue was 485 nm (i.e., the one in between 483 nm and 489 nm), and the “perfect” green was 507 nm (i.e., the one in between 506 nm and 508 nm). For the yellow-to-orange color set, the “perfect” yellow was 582 nm (i.e., the one in between 580 nm and 584 nm), and the “perfect” orange was 604 nm (i.e., the one in between 603 nm and 606 nm). For the orange-to-red color set, the “perfect” orange was 620 nm (i.e., the one in between 619 nm and 622 nm), and the “perfect” red was 643 nm (i.e., the one in between 641 nm and 645 nm). In this manner, the “perfect” color circles identified were located toward the end at the extreme ends of a color set and thus had a high certainty of being judged as a specific color based on the results from the pilot investigation. Each of the three pairs of the “perfect” color circles (i.e., 485 nm versus 507 nm; 582 nm versus 604 nm; and 620 nm versus 643 nm) was printed on the gloss paper in the size of 2.2 inches by 2.2 inches. The resulting three pairs of paper-version “perfect” color circles were inserted into the screens of the two switches respectively as the choices for the participants to select.

Color stimuli used in the experiment. Throughout the experiment, some or all of the color stimuli from each of three color sets validated as a result of the pilot investigation (see Table 11) were used as the samples presented on the laptop. Specifically, during baseline and generalization testing, all nine color stimuli from each of three color sets were used as the samples. During intervention, pre-determined teaching stimulus or stimuli from each color set were used as the samples.

Throughout the experiment (baseline, intervention, and generalization testing), the paper-version “perfect” color pictures that were the same as those used in the pretraining tasks involving color circles were used as choices and placed on the switches for the participants to select. The use of the paper-version “perfect” color circles as the color choices in the experiment ensured that the color choices were different from any of the color samples.

Preference Assessment

A set of preferred items, as identified during individualized preference assessments with each participant (described below), were used as reinforcers during study visits. These items included videos displayed via a portable DVD player and additional edible items (Thomas Train videos for Bruce; Frozen movie video and Lalaloopsy video, and marshmallow and cereal for Gabi; Frozen movie video and Sesame Street videos for Bryan; The Wiggles Dancing videos, and fish crackers for James; and Sesame Street videos for Corbin). For Bruce, playing iPad, social games (e.g., chasing, tag), and playing trains were also used as reinforcers when he became satiated on videos based on the experimenter’s interaction with him. For Corbin, playing hide-and-seek was sometimes used near the end of a visit based on the experimenter’s interaction with him and sometimes based upon his request. Additionally, to minimize the disruption to the task due to a continuous reinforcement (CRF) schedule with tangible reinforcer(s) and thereby enhance the effectiveness of instruction, a penny board was prepared for Bruce, Gabi, and Corbin to enable the delivery of delayed reinforcement. This decision was made in that (a) these three participants understood the concept of the penny board as reported by parents and observed by the experimenters, and (b) it was relatively more effortful and educationally less effective to redirect them back to work if they were provided with tangible reinforcer(s). However, for Bryan and James, the penny board was not used as it proved to less effective in maintaining engagement than continuous reinforcement systems. The board was a 10 cm by 5 cm clipboard, with one white Velcro™ strip pasted in the middle of the board and 10 small white circle Velcro™ spots pasted in parallel with the Velcro™ strip but under it and close to the bottom of the board. Ten pennies were prepared with one of each placed on each of the 10 circle Velcro™ spots. The experimenter moved the pennies from the circle Velcro™ spots to the Velcro™ strip to indicate the forthcoming work, and then moved each penny back to the Velcro™ spot in accordance with the reinforcement schedule required by a specific task. When all pennies on the Velcro™ strip were removed back to the Velcro™ spots, the participants were allowed to have access to reinforcement.

Prior to each assessment involving the responses from participants (as described previously), pretraining, and experimental session, a multiple stimulus without replacement preference assessment (DeLeon & Iwata, 1996) was conducted with each of the participants so as to identify preferred items or activities. The participants were allowed to choose from an array of DVDs (and also edible items for two of the participants [Gabi and James] as suggested by their moms). The item chosen first by each participant was used a reinforcer during a given session, unless the participant indicated preference for a new item during the session by vocally requesting, physically moving toward, or using gestures or signs to request the item (Hahs, 2015).

Research Design

An adapted alternating treatment design (Kazdin, 2010) was replicated across five participants. The simple discrimination condition involved only one teaching stimulus S_A. Participants were taught to respond to one choice but not taught to respond to the remaining choice, with the S_A serving as the sample. In the two conditional discrimination conditions, there were two teaching stimuli (S_A and S_B). Participants were taught to respond to one choice in the presence of the sample S_A and respond to the other choice in the presence of the sample S_B. Across all three conditions, the S_A was always the PSI stimulus of a particular color set. For each participant a specific color set was assigned to only one condition. As Table 12 shows, the assignment of the three color sets into the three experimental conditions was counterbalanced across five participants. In addition, the assignment of the three experimental conditions took the participants' baseline performance into account. For instance, if a participant's baseline data demonstrated that for one color set he or she was able to well discriminate the PSI stimulus and the stimulus with four intervals away from the PSI stimulus, but was not able to well discriminate the PSI stimulus and the stimulus with two intervals away from the PSI, then this color set may be a reasonable candidate to be assigned into the minimal-difference conditional discrimination condition.

Independent variables. The difference between teaching stimuli was the independent variable. The difference between teaching stimuli was operationally defined as the difference of two teaching stimuli (i.e., the S_A [i.e., the PSI stimulus] and the S_B) involved in the conditional discrimination conditions regarding the wavelength and luminosity. As Table 9 shows, in the maximal-difference conditional discrimination condition, the S_B was four intervals away from the S_A in terms of the wavelength and luminosity (about 13 nm difference on wavelength and .12 difference on luminosity), whereas in the minimal-difference one, the S_B was two intervals away from the S_A (about 6.88 nm on wavelength and .06 difference on luminosity). For one participant (i.e., Bruce), the initial difference between the S_A and the S_B was only one interval, which was then increased to two intervals (to be described later). In the simple discrimination condition only the S_A was the teaching stimulus.

Dependent variable. The dependent variable was the percent of independent selection of the target response choice to which the color sample corresponded.

Procedural Overview and General Procedures

Each participant took part in two pretraining tasks prior to receiving the experiment. The first pretraining involved identity matching to sample tasks using familiar pictures, and the second pretraining involved identity matching to sample tasks using "perfect" color circles. The experiment included three experimental conditions and each condition consisted of three phases, including baseline, intervention, and generalization testing. Several general procedures throughout pretraining and experimental sessions were described as follows.

The set-up of materials. Prior to each pretraining or experimental session the laptop was placed on the table at a viewing distance of approximately 25 to 30 inches in front of the participant with the screen of the laptop positioned vertically to the table surface, at the participant's eye level. The two switches were placed on each side of the laptop at a distance of approximately 25 to 30 inches from the participant. In addition, the preference assessment was implemented to identify reinforcer(s). For Bruce, Gabi, and Corbin who used delayed reinforcement, the experimenter placed the identified reinforcer(s) beside the penny board with several blank spots (ranging from 1 to 10) to be filled by pennies, with the number of blank spot(s) determined based on each participant's past performance as reported by parents. For Bryan and James who did not use delayed reinforcement, the experimenter placed the identified tangible reinforcer(s) on the table within their view but out of their reach.

The implementation of one teaching or testing block. Each pretraining or experimental task was comprised of blocks of a specific number of teaching or testing opportunities as preprogrammed by the Presentation® software. Prior to each block, the experimenter directed the participant to look at the paper-version picture choices that had been affixed to each of the switches (e.g., "Look, this is BIRD. This is PIG"). Next, the experimenter loaded the corresponding experimental file in the Presentation® software, entered the participant number, and ran the file. As soon as the file was activated, words saying "READY?" in white letters on a black background were displayed. The experimenter pressed the "Enter" key on the keyboard to start the teaching or testing block after the participant attended to the laptop screen. Then one of the two picture samples was randomly presented on the laptop screen, and the experimenter guided the participant's attention to the screen by pointing and saying like "Look" or "Look at here." Subsequently, the experimenter asked, "Touch the one you see," or "Which one do you see?" or something similar. Participants made a choice selection by pressing one of the switches, and the responses were automatically saved into the logfile of the Presentation® software. Subsequent to a response, a feedback was provided on the screen, which could be a smiley face, an unhappy face, or a blank screen. The next picture was subsequently presented on the screen after a preprogrammed interstimulus interval of 3 s elapsed. When all opportunities within a block were completed, the laptop screen returned to the desktop with the initial display of the Presentation® software.

Prompting procedures during pretraining and intervention tasks. At the beginning of each of pretraining tasks, to make sure the participant understood what he or she was to do, for the first specified number of teaching opportunities the experimenter provided hand-over-hand physical guidance for the participant to touch and click the switch with the correct corresponding picture choice. As soon as the correct picture choice was selected, a big smiley face was shown on the laptop screen. The experimenter said in a happy and enthusiastic voice tone, "Great job, that is [NAME OF THE CORRECT CHOICE]. You got a smiley face." For Bruce, Gabi, and Corbin, the experimenter moved a penny on the Velcro™ strip on the penny board and placed it onto the circle blank Velcro™ spot, and said, "You got a penny." As soon as they got a penny board with full pennies, the experimenter said, "You got all pennies. You can have (the preferred item chosen at the beginning; e.g., iPad)." For Bryan and James, as soon as they produced a correct response, they were allowed access to the preferred item chosen at the beginning. All participants were allowed to have access to the reinforcer for about 10 to 15 s prior to next teaching opportunity.

After the first specified number of opportunities with implementation of the procedures described above that resulted in a correct response, a 3 s constant time delay prompting procedure was implemented. That is, if the participant did not initiate a response within 3 s, the experimenter delivered the least intrusive prompt required for a correct response and the response was recorded as incorrect. For Bruce, Gabi, and Corbin, a combination of a gestural prompt by pointing to the correct switch and a verbal prompt by saying "It is [THE NAME OF THE CORRECT CHOICE]" was implemented. For Bryan and James, the prompt was a combination of a hand-over-hand physical prompt by guiding their hand to the correct switch and a verbal prompt. Subsequent to the prompted correct response, the experimenter provided the feedback, "Ok. It is [THE NAME OF THE CORRECT CHOICE]" but did not provide reinforcement. If the participant produced a correct response within 3 s, a smiley face appeared on the laptop screen, the experimenter said, "Nice job. It is [THE NAME OF THE CORRECT CHOICE]. You got it right" and delivered a penny (for Bruce, Gabi, and Corbin) or the tangible reinforcer (for Bryan and James). After accumulating one to ten pennies as predetermined, the participants (i.e., Bruce, Gabi, and Corbin) exchanged the pennies for the reinforcer.

If the participant produced an incorrect response within 3 s, an unhappy face appeared on the laptop screen, the experimenter said, "It is not [THE NAME OF THE INCORRECT CHOICE]. It is [THE NAME OF THE CORRECT CHOICE]. Nice try" and did not provide a penny (for Bruce, Gabi, and Corbin), or the tangible reinforcer (for Bryan and James). If the participant made incorrect responses during three consecutive opportunities, the experimenter implemented errorless teaching for four opportunities by providing the least amount of prompting for correct response as previously described immediately subsequent to the presentation of a stimulus sample. Subsequently, the 3 s constant time delay procedure was reinstituted until the participant reached the mastery criterion. During this process, reinforcement was provided for every correct independent response but not provided contingent on the prompted correct response or incorrect response.

Feedback and reinforcement procedures. During pretraining and intervention, a big smiley was presented following a correct response and an unhappy was presented following an incorrect response. During the baseline and generalization testing phases of the experiment, a blank screen was presented following a response (i.e., no specific feedback was provided).

During pretraining tasks and the intervention phase of the experiment, the conditioned reinforcer (i.e., the penny) or the immediately delivered tangible reinforcer were provided on a continuous reinforcement schedule, that is, contingent on each correct response. During baseline and generalization testing phases of the experiment, the conditioned reinforcer or the tangible reinforcer were provided on a variable ratio schedule of 3 (VR-3) (i.e., when the participant on average made three responses wherein the correctness of the response was not required). For the three participants to whom the penny board was applied, once the agreed upon number of pennies had been accumulated they were exchanged for a predetermined tangible reinforcer.

Procedures for inappropriate responses. Throughout pretraining and experiment tasks, if the participant attempted to press two switches simultaneously, press another switch after pressing one switch, or press the laptop keyboard, the experimenter physically blocked the attempt by placing his or her hands onto his or her laps, and said, “one press at a time” or “press the switch.” If the experimenter was not quick enough to block him or her, the experimenter still physically prompted his or her hands onto the laps, and said, “one press at a time” or “press the switch.” No reinforcement was provided contingent on these kinds of responses.

In addition, during baseline and generalization testing phases of the experiment, if the participant pressed the two switches at the same time or the experimenter accidentally touched the switch, the response on that opportunity was considered as invalid and was not counted. Table 13 shows the number of the invalid responses during baseline, and generalization testing for each participant. The number of the invalid responses ranged from zero to six across the five participants, with a total of 11 occurrences. During pretraining tasks and the intervention phase of the experiment, if the same situations described above occurred during a teaching opportunity, the response was counted but recorded as incorrect. Finally, all prompted responses during pretraining and intervention phases of the experiment were recorded as incorrect. Next the procedures for the pretraining and the experiment are described respectively.

Pretraining Procedures

Matching to sample pretraining using familiar pictures. Pretraining was designed to ensure that the participants knew how to complete a two-sample two-choice matching to sample task which was the instructional format implemented during the experiment. These sessions utilized familiar pictures to facilitate learning of the response. Four pairs of pictures representing objects or cartoon characters obtained from Google© search engine (i.e., bird versus pig; apple versus pumpkin; Mickey Mouse™ versus Donald Duck™; and Thomas Train™ versus Sponge Bob™) were used as samples presented on the laptop. These pictures were arbitrarily selected and confirmed with the participants' parents to make sure their children were familiar with these objects or cartoon characters. In addition, each of the images was printed on glossy paper in the size of 2.2 inches by 2.2 inches. The resulting four pairs of paper-version objects or cartoon characters were inserted into the screens of the two switches respectively as the choices for the participants to select.

Each teaching block consisted of four teaching opportunities. In each block of four instructional opportunities, each picture sample from one pair of picture samples was presented twice in a random order. After one block of opportunities had been completed, the next block was initiated consisting of a different pair of familiar pictures as the samples and choices. As described previously, errorless prompting, constant time delay prompts, and reinforcement were used throughout this phase to teach the appropriate discriminations. The errorless prompting (i.e., hand-over-hand physical guidance) was implemented for the first block of four teaching opportunities. After the first four opportunities of implementation of the procedure that resulted in a correct response, 3 s constant time delay procedures as described in the general procedures were implemented. The pretraining ended when the participant was able to produce at least 16 correct responses out of 20 consecutive teaching opportunities that spanned five consecutive blocks. All five participants met the criterion.

Matching to sample pretraining on the "perfect" color circles. This pretraining was designed to ensure that participants were able to reliably differentiate between the two "perfect" colors on each color set. The procedures were overall similar to the matching to sample pretraining using pictures of familiar objects or cartoon characters except the use of different samples and choices and the number of teaching opportunities in a block.

As described in the materials section, three pairs of "perfect" color circles (see Table 11) as samples and the corresponding three pairs of "perfect" color circles in the paper version as choices were prepared as the teaching materials. Each block consisted of eight teaching opportunities. In each block, each color sample from one pair of "perfect" colors was presented four times in a random order. After one block of opportunities had been completed, the next block was initiated consisting of the same pair of "perfect" colors as the samples and choices. For each pair of "perfect" colors, the pretraining concluded when the participant produced 14 correct responses during 16 consecutive teaching opportunities that spanned two consecutive blocks. Then the pre-training was initiated for the next pair of "perfect" colors. The pretraining order among the three pairs of "perfect" colors was randomly assigned across five participants.

The teaching procedures were the same as those to those previously described for the matching to sample pretraining involving familiar pictures. One exception was that during the first block of eight teaching opportunities for each pair of "perfect" colors, the experimenter provided hand-over-hand physical guidance for the participant to touch and click the switch with the correct corresponding color picture. After the first eight opportunities of implementation of the procedure that resulted in a correct response, the 3 s constant time delay procedures as described in the general procedures were implemented, until the participant reached the mastery criterion on each pair of “perfect” color circles. All five participants met the criterion.

Experimental Procedures

Following the initial pretraining phases, the experimental phases were implemented. These phases included a baseline phase, followed by an intervention phase that consisted of three concurrent intervention conditions, and a generalization testing phase. Throughout the experiment, color stimuli from each of three color sets validated as a result of the pilot investigation (see Table 11) were used as the samples presented on the laptop. Specifically, during baseline and generalization testing, all nine color stimuli from each of three color sets were used as the samples. During intervention, one or two pre-determined teaching stimuli (depending on the teaching condition) from each color set were used as the samples presented on the computer. The paper-version “perfect” color pictures that were the same as those used in the pretraining tasks involving color circles were used as choices and placed on the switches for the participants to select. The PSI stimulus of each color set was always placed on the participant’s right hand (from the view of the participant facing the laptop screen), and the other choice was placed on the participant’s left hand. Figure 9 presents a demonstration of the placement of “perfect” color circles.

Baseline. During baseline, there were a total of 15 blocks, with five blocks for each color set (5 x 3). Within each block, the nine stimuli from one target color set were randomly presented with each of them presented once, resulting in a total of nine opportunities per block. Thus, there were a total of 135 (15 x 9) testing opportunities. The presentation of these 15 blocks (i.e., five blocks for the blue-to-green color set, five blocks for the yellow-to-orange color set, and five blocks for the orange-to-red color set) was randomly alternated. The baseline testing was completed in one session for Bruce, Gabi, Bryan, and Corbin, and in two sessions for James.

Within each block of nine opportunities, as a color sample was displayed, the experimenter directed the participant's attention to the laptop screen saying, "look," and delivered the instruction "Is it blue or green?" or "What color do you see?" If a participant did not respond within 30 s of a verbal directive, the directive was repeated. Subsequent to a response the next color sample was displayed paired with the same instruction. Reinforcement for each participant was provided on a variable ratio schedule of 3 (VR-3) and was contingent only on learner participation. For Bruce, Gabi, and Corbin, a penny board as described earlier was implemented to signal the delivery of the reinforcer[s]. Tangible reinforcers were delivered for Bryan and James.

For each participant, the baseline generalization gradient for each color set was examined to ensure it met the following two criteria, (a) it was not exactly the same as the ideal generalization gradient (see Figure 8), in which the percent of independent selection of a target choice was 100% given the PSI stimulus and the stimuli from stimulus Class A, and it was 0% given the stimuli from stimulus Class B. Recall that for the stimuli within stimulus Class A, the correct response choice was the same as the one to which the PSI stimulus corresponded. For the stimuli within stimulus Class B, the correct response choice was the other option, and (b) the RSS value of the baseline generalization gradient was larger than that obtained from typically developing adults on the corresponding color set. Recall that the RSS value of the generalization gradient obtained from typically developing adults in the pilot study was .11 for the blue-to-green color set, .64 for the yellow-to-orange color, and .42 for the orange-to-red color set. None of the five participants' baseline data were the same as the ideal generalization gradient, thus meeting the criterion (a). The RSS values of their baseline generalization gradients as shown in Table 14 demonstrated that all five participants also met the criterion (b). Therefore, all their baseline data indicated sufficient room for improvement. On the basis of the baseline performance, the three color sets were then assigned into the three intervention conditions and counterbalanced across five participants.

Intervention. Throughout the intervention for each experimental condition, each block of instructional opportunities consisted of 10 teaching opportunities. The three experimental conditions were conducted in an alternated manner within each session. The mean number of teaching opportunities per session ranged from 51 to 102 teaching opportunities across the five participants.

Each block of 10 opportunities began with the presentation of a stimulus sample. As soon as the color sample was displayed, the interventionist directed the participant's attention to the screen, and delivered instructional directive like, "Is it 'Blue' or 'Green'?" in accordance with the assigned color set, while pointing to the two paper-version choices placed on the two switches respectively. The next teaching opportunity did not start until the participant had produced a response. Prompting, error correction, and reinforcement procedures were implemented as described in the general procedures (p. 74 to p. 79) until the participants reached the mastery criterion on each condition.

Simple discrimination condition. The S_A was only the teaching stimulus in this condition. As a result, the same response was correct on all teaching opportunities during this condition for each learner. Each block of 10 opportunities involved the presentation of the stimulus sample (i.e., the PSI stimulus from the color set assigned as the simple discrimination condition). A 3 s constant time delay prompting procedure as described in the general procedures was implemented during instruction. The mastery criterion for the simple discrimination condition was at least 80% correct independent responses across two consecutive blocks of 10 opportunities.

Maximal-difference conditional discrimination condition. During the maximal-difference conditional discrimination condition, one intervention block consisted of 10 teaching opportunities (5 opportunities for one teaching stimulus sample S_A [i.e., the PSI stimulus], 5 opportunities for the other teaching stimulus sample S_B [i.e., the stimulus most different from the S_A, with four-interval difference between two stimuli]). The intervention procedures were the same as those implemented in teaching a simple discrimination with one exception regarding the mastery criterion. Mastery criterion for the maximal-difference conditional discrimination training was at least 80% correct in the presence of S_A and S_B in one block (i.e., 4 out 5 correct for each color sample) and across two consecutive blocks of 10 opportunities.

Minimal-difference conditional discrimination condition. During the minimal-difference conditional discrimination condition, one intervention block consisted of 10 teaching opportunities (5 opportunities for one teaching stimulus sample S_A [i.e., the PSI stimulus], 5 opportunities for the other teaching stimulus sample S_B [i.e., the stimulus in between the PSI stimulus and the one most different from the S_A, with two-interval difference between two stimuli). The difference between S_A and S_B was empirically validated for Bruce who was the first to participate in the study. For Bruce the initial minimal-difference conditional discrimination involved only one-interval difference between S_A and S_B. His learning performance in this condition demonstrated little improvement across intervention blocks. Therefore, after a return to the baseline (i.e., a re-baseline), the conditional discrimination intervention was implemented with the difference between S_A and S_B going up to two intervals. The intervention procedures and mastery criterion were the same as those described for the maximal-difference conditional discrimination.

Generalization testing. As soon as a participant reached the mastery criterion on one color set under an experimental condition, generalization testing was initiated for the participant and for that color set. Implementation procedures were the same as during baseline phase. Generalization testing for each color set consisted of five blocks, resulting in a total of 45 opportunities for each color set that was assigned into a specific experimental condition.

Interobserver Agreement

Response Reliability. Inter-observer agreement (IOA) between the experimenter and a second independent observer was assessed during a minimum of 35% blocks in each phase (i.e., baseline, intervention, and generalization testing) for all five participants. IOA for each block was calculated as dividing the opportunities of agreements by the total opportunities of agreements and disagreements, and converting the resulting quotient into a percent. For baseline and generalization testing, an agreement was defined as both the experimenter and the independent observer recording the participant pressed the switch on one specific side. For intervention, an agreement was defined as both the experimenter and the independent observer recording the participant's response as correct or incorrect.

Table 15 shows the results of the response reliability during each phase of the experiment for each participant, as well as the number of blocks in which the IOA was taken. Briefly, the IOA was collected during 100% of all intervention and testing blocks for Bruce, 67.27% of all blocks for Gabi, 48.39% of all blocks for Bryan, 54.46% of all blocks for James, and 88.14% of all blocks for Corbin. IOA was 100% for all participants. In addition, all data were doubled checked with the corresponding log files automatically produced by the Presentation® program, and all correct responses recorded by the experimenter were also recorded as correct in the log files.

Procedural fidelity. Procedural fidelity involved task analyzing each step of the baseline, intervention, and generalization testing procedures being implemented. Using an experimenter-developed checklist, an independent in-person observer independently checked procedural fidelity for each testing or teaching opportunity among at least 35% of the blocks across baseline, intervention, and generalization testing for each participant. The checklist consisted of the following procedural steps, which included: (a) securing the participant's attention, (b) using the correct stimulus, (c) delivering the correct question, (d) providing the appropriate delay interval between instruction and the participant's response, (e) delivering the controlling prompt (if needed), (f) providing contingent positive reinforcement correctly, (g) delivering instructional feedback during the intervention phase, and (h) not delivering the instructional feedback during baseline and generalization testing phases. The IOA was calculated for each block as dividing the opportunities of agreements by the total opportunities of agreements and disagreements, and multiplied by 100.

Table 16 shows the results of the procedural fidelity across phases for each participant, as well as the number of blocks in which the IOA was taken. Specifically, procedural fidelity data were taken during 100% of all blocks for Bruce, 67.27% of all blocks for Gabi, 48.39% of all blocks for Bryan, 54.46% of all blocks for James, and 77.97% of all blocks for Corbin. The average IOA for procedural fidelity was 96.59% for Bruce (range: 86.67% - 100%), 98.96% for Gabi (range: 93.33% - 100%), 97.66% for Bryan (range: 93.33% - 100%), 99.72% for James (range: 97.78% - 100%), and 99.46% for Corbin (range: 97.78% - 100%).

Data Analysis Procedures

To answer the first research question regarding intervention efficiency, the number of teaching opportunities to mastery criterion for each experimental condition was calculated. To answer the second research question regarding generalization performance, the results were analyzed using visual inspection, with the guidelines established by Kratochwill and colleagues (2010). Conventionally, visual inspection involved analysis of changes in level, trend, and variability between adjacent phases to determine whether an experimental effect was evident. For this experiment, these guidelines were used to compare the adjacent generalization gradients, that is, comparing the baseline generalization gradient and the generalization gradient obtained subsequent to intervention.

No effect size metrics (e.g., PND, NAP, PAND) were calculated in this study since there has been little evidence on how to compute an effect size for generalization gradients in a single case experimental design.

CHAPTER IV: RESULTS

Figures 10(a) to 10(e) display the results during baseline, intervention, and generalization testing phases for each participant. Each figure consists of three panels, corresponding to the performance on the three conditions (i.e., simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions). For generalization gradient data, the abscissa indicates the stimulus values of color stimuli (in wavelength and luminosity value); whereas the ordinate describes the percent of independent selections of the choice given the sample S_A (i.e., the PSI stimulus), and the percent of independent selection of the other choice is its exact complement. For intervention data, the abscissa indicates the intervention blocks of opportunities for a specific teaching stimulus; whereas the ordinate describes the percent of independent selection of the target choice to which each teaching stimulus corresponded out of the total presentation opportunities of each teaching stimulus in one intervention block.

Table 17 shows the percent of responses to the S_A (i.e., the PSI stimulus), the mean percent of responses to the four stimuli within the stimulus Class B (i.e., the stimuli that were categorized as not sharing the color category assigned to the PSI stimulus for teaching purposes), and the mean percent of responses to the four stimuli within the stimulus Class A (i.e., the stimuli that were categorized as sharing the color category assigned to the PSI stimulus for the teaching purposes) for each generalization gradient during baseline and generalization testing. Recall that for the stimuli within stimulus Class A, the correct response choice should be the one to which the PSI stimulus corresponded. For the stimuli within stimulus Class B, the correct response choice should be the other option. Results that follow are organized individually for Bruce, Gabi, Bryan, James, and Corbin, and finally grouped together across five participants for a participant comparison.

Bruce

Simple Discrimination Condition

The first panel of Figure 10(a) shows Bruce's performance across each phase on the simple discrimination condition (assigned with the orange-to-red color set).

Baseline. In Figure 10(a), the far left graph of the first panel shows that during baseline, Bruce selected the response choice ("Red") that corresponded to the sample S_A (i.e., the PSI stimulus of the orange-to-red color set) at similar rates across all nine color samples that were made available. Consequently, the generalization gradient did not resemble the ideal sigmoidal form. He selected the "Red" choice when presented with the sample S_A of 631 nm during 40% of opportunities. He selected the "Red" choice a mean of 55% of opportunities when presented with the sample stimuli from the orange stimulus class (i.e., 619, 622, 625, 628 nm), and a mean of 50% of opportunities when presented with the sample stimuli from the red stimulus class (i.e., 634, 637, 641, 645 nm). There was almost no differentiation in using the classifiers orange and red.

Intervention. In Figure 10(a), the far right graph of the first panel shows Bruce required 30 teaching opportunities to master the simple discrimination involving the selection of the "Red" choice in the presence of the sample S_A of 631 nm.

Generalization testing. In Figure 10(a), the center graph of the first panel shows that subsequent to intervention, Bruce's performance in the implementation of the generalization gradient assessment more closely approximated a sigmoidal form. The percent of selection of "Red" choice at the sample S_A of 631 nm increased from 40% during baseline to 80%. Also, the responding to the red and orange classifiers was fairly differentiated. The mean percent of the "Red" choice for the sample stimuli within the orange stimulus class was 31.25%, which was a decrease from 55% during baseline. The mean percent of the "Red" choice for the sample stimuli within the red stimulus class was 68.75%, which was an increase from 50% during baseline.

There was a difference of 37.5% in the mean percent of response to the two stimulus classes, which was an increase from -5% during baseline.

Maximal-Difference Conditional Discrimination Condition

The second panel of Figure 10(a) shows Bruce's performance on the maximal-difference conditional discrimination condition that was assigned with the yellow-to-orange color set.

Baseline. In Figure 10(a), the far left graph of the second panel shows that the baseline generalization gradient was not quite in a sigmoidal form; especially the portion of the generalization gradient associated with the yellow stimulus class (i.e., 580, 584, 588, 591 nm). The performance within the yellow stimulus class seemed to be monotonically increasing, whereas the portion of the generalization gradient for the sample stimuli associated with the orange stimulus class (i.e., 597, 600, 603, 606 nm) demonstrated greater fluctuations. He selected the "Orange" choice when presented with the PSI stimulus of 594 nm during 60% of opportunities. He selected the "Orange" choice a mean of 28.75% of opportunities when presented with the sample stimuli from the yellow stimulus class, and a mean of 75% of opportunities when presented with the sample stimuli from the orange stimulus class. There was a difference of 46.25% in the mean response to the two stimulus classes.

Intervention. In Figure 10(a), the far right graph of the second panel shows that Bruce required 60 teaching opportunities to master the maximal-difference conditional discrimination between the sample S_A of 594 nm (representing orange) and the sample S_B of 580 nm (representing yellow).

Generalization testing. In Figure 10(a), the center graph of the second panel shows that subsequent to intervention Bruce's performance demonstrated a response bias toward the "Orange" choice, with an increase of the "Orange" responses to both stimulus classes. The percent selection of the "Orange" choice given the sample S_A increased to 80% compared to the 60% observed during baseline. He selected the "Orange" choice when presented with sample stimuli associated with the orange stimulus class during a mean of 85% of opportunities, which was an increase from 75% during baseline. He selected the “Orange” choice when presented with sample stimuli associated with the yellow stimulus class during a mean of 40% of opportunities, which was also an increase from 28.75% during baseline (i.e., an indication of over-generalization of the “Orange” choice). Taken together there was a mean difference of 45% between performance in the two stimulus classes, which was a slight decrease from 46.25% during baseline.

Minimal-Difference Conditional Discrimination Condition

The third panel of Figure 10(a) shows Bruce’s performance on the minimal-difference conditional discrimination condition assigned with the blue-to-green color set.

Baseline. In Figure 10(a), the first graph of the third panel shows that during baseline Bruce’s responding was variable within both the blue stimulus class and the green stimulus class. The baseline generalization gradient did not quite approximate the ideal sigmoidal form. He selected the “Green” choice given the sample S_A during 80% of opportunities. He selected the “Green” choice a mean of 20% of opportunities when presented with sample stimuli associated with the blue stimulus class (i.e., 483, 489, 493, 496 nm). He selected the “Green” choice a mean of 62.5% of opportunities when presented with sample stimuli associated with the green stimulus class (i.e., 502, 504, 506, 508 nm). There was a 42.5% difference in the mean percent of response to the two stimulus classes.

Initial intervention. In Figure 10(a), the fourth graph of the third panel shows that during the initial intervention involving the discrimination between the sample S_A of 499 nm (representing green) and the sample S_B of 496 nm (representing blue) as one interval, Bruce had not reached the mastery criterion and consistently did not show any improvement after 150 teaching opportunities.

Re-baseline. In Figure 10(a), the second graph of the third panel shows that during the re-baseline, Bruce's performance did not change appreciably from the initial baseline. The response variability remained within both stimulus classes Blue and Green. The percent of "Green" choices given the sample S_A of 499 nm decreased from 80% during baseline to 60%. Bruce selected the "Green" choice at a mean of 15% of opportunities when presented with sample stimuli associated with the blue stimulus class, and at a mean of 65% of opportunities when presented with sample stimuli associated with the green stimulus class. There was a 50% difference in the mean percent of response to the two stimulus classes, comparable to that during the original baseline.

Modified intervention. Bruce had not improved after 150 teaching opportunities during intervention. Consequently, following a re-baseline, a modified intervention with the difference between two teaching stimuli going up from one interval to two intervals. As the fourth graph on the third panel of Figure 10(a) shows, during the modified intervention, Bruce required 120 teaching opportunities to master the conditional discrimination between the sample S_A of 499 nm (representing green) and the new sample S_B of 493 nm (representing blue).

Generalization testing. In Figure 10(a), the third graph of the third panel shows that following mastery, a generalization gradient assessment was reimplemented. Bruce's performance improved to more closely approximate the ideal sigmoidal form. Although the percent of "Green" choice given the sample S_A of 499 nm (representing green) slightly decreased from 60% from re-baseline to 50%, his performance within both stimulus classes became less variable. The mean percent of the "Green" choice was 10% for sample stimuli within the blue stimulus class, which was a decrease from 15% during baseline. The mean percent of the "Green" choice was 95% for sample stimuli within the green stimulus class, which was an increase from 65% during baseline. There was a difference of 85% in the mean response to the two stimulus classes, which was an increase from 50% during baseline.

Comparison Across Experimental Conditions

Bruce required the least number of teaching opportunities (30) to master the simple discrimination with the maximal-difference condition being mastered with 60 teaching opportunities. The minimal-difference conditional discrimination required the greatest number of teaching opportunities (120) to reach mastery. This is consistent with the predicted order across experimental conditions given the available literature (Eckerman, 1970; Galizio, 1980; LaBerge, 1961).

From baseline to generalization testing, both the minimal-difference conditional discrimination and simple discrimination conditions resulted in reasonably equivalent improvement in performance within each of the two stimulus classes, as well as the overall performance. The simple discrimination was most effective in increasing the percent of selection of the target choice given the sample S_A. The maximal-difference conditional discrimination did not improve the overall performance or the performance within the stimulus Class B (i.e., the yellow stimulus class in this scenario), although it improved the performance within the stimulus Class A (i.e., the orange stimulus class in this scenario).

Gabi

Simple Discrimination Condition

The first panel of Figure 10(b) shows Gabi's performance on the simple discrimination condition (assigned with the orange-to-red color set).

Baseline. In Figure 10(b), the far left graph of the first panel shows that during baseline, Gabi's performance did not approximate the ideal sigmoidal form. She selected the "Red" choice given the sample S_A (i.e., 631 nm) during 40% of opportunities. Her performance between the orange stimulus class (i.e., 619, 622, 625, 628 nm) and the red stimulus class (i.e., 634, 637, 641, 645 nm) was differentiated, although her performance in the orange stimulus class was somewhat variable. The mean percent of the "Red" choice was 25% within the orange stimulus class, and 85% within the red stimulus class. There was a difference of 60% in the mean percent of response to the two stimulus classes.

Intervention. In Figure 10(b), the far right graph of the first panel shows that Gabi required 60 teaching opportunities to master the simple discrimination involving the selection of "Red" choice at the presentations of the sample S_A (i.e., 631 nm).

Generalization testing. In Figure 10(b), the center graph of the first panel shows that subsequent to intervention, Gabi's overall performance demonstrated both improvement and deterioration. A response bias toward the "Red" choice was observed, with an increase of the "Red" responses to both stimulus classes. The percent of "Red" choice given the sample S_A of 631 nm increased from 40% to the ideal 100%. The mean percent of "Red" choice was 35% within the orange class, which was a slight increase from 25% during baseline (i.e., an indication of over-generalization of the "Red" response). Her performance given the sample stimuli from the red stimulus class somewhat improved. The mean percent of "Red" choice was 95% within the red class, which was a slight increase from 85% during baseline. The difference in the mean responding to the two stimulus classes remained at 60%, the same as that during baseline.

Maximal-Difference Conditional Discrimination Condition

The second panel of Figure 10(b) presents Gabi's performance on the maximal-difference conditional discrimination condition (assigned with the blue-to-green color set).

Baseline. In Figure 10(b), the far left graph of the second panel shows that during baseline Gabi's performance within the blue stimulus class deviated from the ideal performance and demonstrated greater variability than her performance within the green stimulus class. She selected the "Green" choice given the sample S_A (i.e., 499 nm) during 80% of opportunities. She selected the "Green" choice at a mean of 40% of opportunities when presented with sample stimuli associated with the blue stimulus class (i.e., 483, 489, 493, 496 nm), and at a mean of 95% of opportunities when presented with sample stimuli associated with the green stimulus class (i.e., 502, 504, 506, 508 nm). The difference between the mean percent of response to the two stimulus classes was 55%.

Intervention. In Figure 10(b), the far right graph of the second panel shows that Gabi needed 30 teaching opportunities to master the maximal-difference conditional discrimination between the sample S_A of 499 nm (representing green) and the sample S_B of 483 nm (representing blue).

Generalization testing. In Figure 10(b), the center graph of the second panel shows that following mastery Gabi's performance during the implementation of the generalization gradient assessment became closer to the ideal sigmoidal form. The percent of selection of "Green" choice given the sample S_A of 499 nm was 80%, which was the same as that during baseline. The mean percent of "Green" choice was 20% within the blue stimulus class, which was a decrease from 40% during baseline. The mean percent of "Green" choice was 95% within the green stimulus class, the same as that during baseline. There was a 75% difference in the mean percent of response to the two stimulus classes, which was an increase from 55% during baseline.

Minimal-Difference Conditional Discrimination Condition

The third panel of Figure 10(b) shows Gabi's performance on the minimal-difference conditional discrimination condition (assigned with the yellow-to-orange color set).

Baseline. In Figure 10(b), the far left graph of the third panel shows that during baseline Gabi's performance did not approximate an ideal sigmoidal form. Particularly, a clear fluctuation was observed in her response within the orange stimulus class (i.e., 597, 600, 603, 606 nm). She selected the "Orange" choice in the presence of the sample S_A (i.e., 594 nm) during 100% of opportunities. The mean percent of "Orange" choice was 20% for sample stimuli within the yellow stimulus class (i.e., 580, 584, 588, 591 nm), and 75% for sample stimuli within the orange stimulus class. There was a 55% difference in the mean percent of response to the two stimulus classes.

Intervention. In Figure 10(b), the far right graph of the third panel shows that Gabi needed

160 teaching opportunities to master the minimal-difference conditional discrimination between the sample S_A of 594 nm (representing orange) and the sample S_B of 588 nm (representing yellow).

Generalization testing. In Figure 10(b), the center graph of the third panel shows that following mastery, Gabi's performance during the implementation of the generalization gradient assessment more closely approximated the ideal sigmoidal form. The percent of "Orange" choice in the presence of the sample S_A of 594 nm declined from 100% during baseline to 60%. The mean percent of "Orange" choice was 15% for sample stimuli within the yellow stimulus class, which was a slight decrease from 20% during baseline. The mean percent of "Orange" choice was 85% for sample stimuli within the orange stimulus class, which was an increase from 75% during baseline. There was a 70% difference in the mean percent of response to the two stimulus classes, which was an increase from 55% during baseline.

Comparison Across Experimental Conditions

Gabi required the least number of teaching opportunities (30) to master the maximal-difference conditional discrimination, with the simple discrimination being mastered with 60 teaching opportunities. The minimal-difference conditional discrimination required the greatest number of teaching opportunities (160) to reach mastery. This is inconsistent with the predicted order given the existing literature (Eckerman, 1970; Galizio, 1980; LaBerge, 1961).

From baseline to generalization testing, the two conditional discrimination conditions were relatively more effective in improving the overall performance than the simple discrimination condition. The simple discrimination enhanced the correct responding to the sample S_A (i.e., the sample stimulus of 631 nm representing red in this scenario), whereas the two conditional discriminations did not enhance the correct responding to the corresponding sample S_A. All three conditions improved the performance within corresponding stimulus Class A. The two conditional discrimination conditions were more effective in enhancing the performance within corresponding stimulus Class B than the simple discrimination condition.

Bryan

Simple Discrimination Condition

The first panel of Figure 10(c) shows Bryan's performance on the simple discrimination condition (assigned with the yellow-to-orange color set).

Baseline. In Figure 10(c), the far left graph of the first panel shows that during baseline the generalization gradient was far away from the ideal sigmoidal form. Bryan's performance was somewhat variable across all nine color stimuli and the differentiation in his responding to the two stimulus classes was in a direction opposite to the ideal one. He selected the "Orange" choice given the sample S_A (i.e., 594 nm) on 20% of opportunities. The mean percent of "Orange" choice was 45% over the sample stimuli from the yellow stimulus class (i.e., 580, 584, 588, 591 nm), and 40% over the sample stimuli from the orange stimulus class (i.e., 597, 600, 603, 606 nm). The difference in the mean percent of response to the two stimulus classes was -5%.

Intervention. In Figure 10(c), the far right graph of the first panel shows that Bryan required 160 teaching opportunities to master the simple discrimination involving the selection of "Orange" choice given the sample S_A (i.e., 594 nm).

Generalization testing. In Figure 10(c), the center graph of the first panel shows that subsequent to intervention, Bryan's performance during the implementation of the generalization gradient assessment improved compared to the baseline. The generalization gradient more closely approximated the ideal sigmoidal form. The responding to the two stimulus classes was more differentiated and the differentiation became in a direction consistent with the ideal one. The percent of "Orange" choice in the presence of the sample S_A (i.e., 594 nm) remained at 20%. The mean percent of "Orange" choice was 40% for sample stimuli within the yellow stimulus class, which was a slight decrease from 45% during baseline. The mean percent of "Orange" choice was

80% for sample stimuli within the orange stimulus class, which was an increase from 40% during baseline. There was a 40% difference in the mean percent of response to the two stimulus classes, which was an increase from -5% during baseline.

Maximal-Difference Conditional Discrimination Condition

The second panel of Figure 10(c) depicts Bryan's performance on the maximal-difference conditional discrimination condition (assigned with the orange-to-red color set).

Baseline. In Figure 10(c), the far left graph of the second panel shows that during baseline the generalization gradient was far away from the ideal sigmoidal form. Bryan demonstrated considerable variability in the responding to the nine sample stimuli, and the differentiation in his responding to the two stimulus classes seemed to be in a direction opposite to the ideal one. The percent of "Red" choice in the presence of the sample S_A (i.e., 631 nm) was 0%. The mean percent of "Red" choice was 55% for sample stimuli within the orange stimulus class (i.e., 619, 622, 625, 628 nm), and 35% for sample stimuli within the red stimulus class (i.e., 634, 637, 641, 645 nm). Thus, the differentiation in the responding to the two stimulus classes was -20%.

Intervention. In Figure 10(c), the far right graph of the second panel shows that Bryan required 470 teaching opportunities to master the maximal-difference conditional discrimination between the sample S_A of 631 nm (representing red), and the sample S_B of 619 nm (representing orange).

Generalization testing. In Figure 10(c), the center graph of the second panel shows that following mastery Bryan's performance during the implementation of the generalization gradient assessment more closely approximated the ideal sigmoidal form. His responding to the two stimulus classes was more clearly differentiated and the differentiation became in a direction consistent with the ideal one. The percent of "Red" choice in the presence of the sample S_A (i.e., 631 nm) increased from 0% during baseline to 60%. He selected the "Red" choice at a mean of 30% of opportunities when presented with sample stimuli within the orange stimulus class, which was a decrease from 55% during baseline. He selected the “Red” choice at a mean of 90% of opportunities when presented with sample stimuli within the red stimulus class, which was an increase from 35% during baseline. There was a difference of 60% in the mean percent of response to the two stimulus classes, which was an increase from -20% during baseline.

Minimal-Difference Conditional Discrimination Condition

The third panel of Figure 10(c) shows Bryan’s performance on the minimal-difference conditional discrimination condition (assigned with the blue-to-green color set).

Baseline. In Figure 10(c), the far left graph of the third panel shows that during baseline similar to Bryan’s performance on the other two conditions, his responses demonstrated clear variability across the nine color stimuli. The baseline generalization gradient was far away from the ideal sigmoidal form and it demonstrated a trend opposite to the ideal one. He selected the “Green” choice in the presence of the sample S_A (i.e., 499 nm) during 60% of opportunities. The mean percent of “Green” choice was 50% for sample stimuli within the blue stimulus class (i.e., 483, 489, 493, 496 nm), and 45% for sample stimuli within the green stimulus class (i.e., 502, 504, 506, 508 nm). Thus, the differentiation in the responding to the two stimulus classes was not only minimal (5%) but also in a direction opposite to the ideal one.

Intervention. In Figure 10(c), the far right graph of the third panel shows that Bryan required 310 teaching opportunities to master the minimal-difference conditional discrimination between the sample S_A of 499 nm (representing green) and the sample S_B of 493 nm (representing blue).

Generalization testing. In Figure 10(c), the center graph of the third panel shows that subsequent to intervention, Bryan’s performance during the implementation of generalization gradient assessment improved and the generalization gradient approximated the ideal sigmoidal form. He selected the “Green” choice in the presence of the sample S_A (i.e., 499 nm) during 80% of opportunities. The mean percent of “Green” choice was 21.25% for sample stimuli within the blue stimulus class, which was a decrease from 50% during baseline. The mean percent of "Green" choice was 100% for sample stimuli within the green stimulus class, which was an appreciable increase from 45% during baseline. There was a large difference of 78.75% in the responding to the two stimulus classes, which was an increase from -5% during baseline.

Comparison Across Experimental Conditions

Bryan required the least number of teaching opportunities (160) to master the simple discrimination, with the minimal-difference conditional discrimination being mastered with 310 teaching opportunities. The maximal-difference conditional discrimination required the greatest number of teaching opportunities (470) to reach mastery. This is inconsistent with the predicted order based on the available literature (Eckerman, 1970; Galizio, 1980; LaBerge, 1961).

From baseline to generalization testing all three conditions improved the overall performance. The two conditional discriminations were effective in enhancing the correct responses to the corresponding sample S_A, whereas the simple discrimination did not enhance the correct response to the sample S_A (i.e., 594 nm). All three conditions enhanced the performance within corresponding stimulus Class A. The two conditional discrimination conditions (especially the minimal-difference one) were more effective in enhancing the performance within corresponding stimulus Class B than the simple discrimination condition. Although the simple discrimination was effective in enhancing the performance within stimulus Class A (i.e., the orange stimulus class in this scenario), it only slightly improved the performance within stimulus Class B (i.e., the yellow stimulus class in this scenario).

James

Simple Discrimination Condition

The first panel of Figure 10(d) shows James' performance on the simple discrimination condition (assigned with the yellow-to-orange color set).

Baseline. In Figure 10(d), the far left graph of the first panel shows that during baseline James' responding to the nine stimuli was variable and the generalization gradient was far away from the ideal sigmoidal form. He selected the "Orange" choice in the presence of the sample S_A (i.e., 594 nm) during 25% of opportunities. The mean percent of "Orange" choice was 42.5% for sample stimuli within the yellow stimulus class (i.e., 580, 584, 588, 591 nm), and 40% for sample stimuli within the orange stimulus class (i.e., 597, 600, 603, 606 nm). The difference between the mean percent of response to the two stimulus classes was -2.5%, which was not only minimal but also in a direction opposite to the ideal one.

Intervention. In Figure 10(d), the far right graph of the first panel shows that James required 90 teaching opportunities to master the simple discrimination involving the selection of "Orange" choice in the presence of the sample S_A (i.e., 594 nm).

Generalization testing. In Figure 10(d), the center graph of the first panel shows that subsequent to intervention, James' performance during the implementation of generalization gradient assessment allocated a majority of his responses to the "Orange" choice across most of the nine color stimuli (i.e., a response bias toward the "Orange" choice). His performance given the sample S_A of 594 nm and the sample stimuli within the orange stimulus class approximated the ideal performance. He selected the "Orange" choice in the presence of the sample S_A (i.e., 594 nm) during 80% of opportunities, which was an increase from 25% during baseline. The mean percent of "Orange" choice was 85% within the orange stimulus class, which was an increase from 40% during baseline. His performance within the yellow stimulus class still deviated from the ideal performance. The mean percent of "Orange" choice was 65% within the yellow stimulus class, which was an increase from 42.5% during baseline. This was an indication of over-generalization of the "Orange" choice. The difference between the mean percent of response to the two stimulus classes was 20%, which was an increase from -2.5% during baseline.

Maximal-Difference Conditional Discrimination Condition

The second panel of Figure 10(d) shows James' performance on the maximal-difference conditional discrimination condition (assigned with the blue-to-green color set).

Baseline. In Figure 10(d), the far left graph of the second panel shows that during baseline James' performance was quite variable across the nine color stimuli. The generalization gradient was far away from the ideal sigmoidal form. He selected the "Green" choice in the presence of the sample S_A (i.e., 499 nm) during 60% of opportunities. The percent of "Green" choice was 40% for sample stimuli within the blue stimulus class (i.e., 483, 489, 493, 496 nm), and 50% for sample stimuli within the green stimulus class (i.e., 502, 504, 506, 508 nm). There was a 10% difference in the mean percent of response to the two stimulus classes.

Intervention. In Figure 10(d), the far right graph of the second panel shows that James needed 420 teaching opportunities to master the maximal-difference conditional discrimination between the sample S_A of 499 nm (representing green) and the sample S_B of 483 nm (representing blue).

Generalization testing. In Figure 10(d), the center graph of the second panel shows that following mastery James' performance during the implementation of generalization gradient assessment did not improve compared to the baseline. His responses over the nine color stimuli were still variable, as reflected by the fluctuations of the generalization gradient. He selected the "Green" choice given the sample S_A of 499 nm during 20% of opportunities, which was a decrease from 60% during baseline. The mean percent of "Green" choice was 50% for sample stimuli within the blue stimulus class, which was an increase from 40% during baseline. This was an indication of over-generalization of the "Green" choice. The mean percent of "Green" choice was also 50% for sample stimuli within the green stimulus class, which was the same as that during baseline. Thus, no differentiation was observed in the mean percent of response to the two stimulus classes, which was a decrease from 10% during baseline.

Minimal-Difference Conditional Discrimination Condition

The third panel of Figure 10(d) shows James' performance on the minimal-difference conditional discrimination condition (assigned with the orange-to-red color set).

Baseline. In Figure 10(d), the far left graph of the third panel shows that during baseline James' performance was far away from the ideal sigmoidal form. The general trend of the generalization gradient was in a direction opposite to the ideal one. He selected the "Red" choice in the presence of the sample S_A (i.e., 631 nm) during 80% of opportunities. The mean percent of "Red" choice was 73.75% for sample stimuli within the orange stimulus class (i.e., 619, 622, 625, 628 nm), and 55% for sample stimuli within the red stimulus class (i.e., 634, 637, 641, 645 nm). The difference in the mean percent of response to the two stimulus classes was -18.75%, which was in a direction opposite to the ideal one.

Intervention. In Figure 10(d), the far right graph of the third panel shows that James required 200 teaching opportunities to master the minimal-difference conditional discrimination between the sample S_A of 631 nm (representing red) and the sample S_B of 625 nm (representing orange).

Generalization testing. In Figure 10(d), the center graph of the third panel shows that following mastery James' performance during the implementation of generalization gradient assessment did not improve toward the ideal sigmoidal form compared to the baseline. The general trend of the generalization gradient was still in a direction opposite to the ideal one. In general his performance demonstrated both improvement and deterioration. A response bias toward the "Orange" choice was observed, with a decrease of the "Red" responses to both stimulus classes. He selected the "Red" choice in the presence of the sample S_A (i.e., 631 nm) during 60% of opportunities, which was a decrease from 80% during baseline. The mean percent of "Red" choice was 60% for sample stimuli associated with the orange stimulus class, which was a decrease from 73.75% during baseline. The mean percent of "Red" choice was 40% for sample stimuli associated with the red stimulus class, which was a decrease from 55% during baseline. This was an indication of under-generalization of the "Red" choice or an indication of over-generalization of the "Orange" choice. The difference in the mean percent of response to the two stimulus classes was -20%, which was a slight decrease from -18.75% during baseline.

Comparison Across Experimental Conditions

James required the least number of teaching opportunities (90) to master the simple discrimination with the minimal-difference conditional discrimination being mastered with 200 teaching opportunities. The maximal-difference conditional discrimination required the greatest number of teaching opportunities (420) to reach mastery. This is inconsistent with the predicted order given the existing literature (Eckerman, 1970; Galizio, 1980; LaBerge, 1961).

From baseline to generalization testing, none of the three conditions improved the overall performance toward the ideal generalization gradient of a sigmoidal form. However, the simple discrimination condition was relatively more effective in enhancing the overall performance toward the ideal generalization gradient of a sigmoidal form than the two conditional discriminations. The simple discrimination was most effective in enhancing the correct responses to the sample S_A and the performance within stimulus Class A (i.e., the orange stimulus class in this scenario). The two conditional discriminations did not enhance the correct responses to the sample S_A and the performance within stimulus Class A. None of the three conditions effectively improved the performance within corresponding stimulus Class B.

Corbin

Simple Discrimination Condition

The first panel of Figure 10(e) shows Corbin's performance on the simple discrimination condition (assigned with the blue-to-green color set).

Baseline. In Figure 10(e), the far left graph of the first panel shows that during baseline

Corbin's performance was close to the ideal sigmoidal form. However, his responses to the sample S_A (i.e., 499 nm) and the sample stimuli associated with the green stimulus class (i.e., 502, 504, 506, 508 nm) were not completely in the ideal form. He selected the "Green" choice given the sample S_A of 499 nm during 40% of opportunities. The mean percent of "Green" choice was 0% for sample stimuli within the blue stimulus class (i.e., 483, 489, 493, and 496 nm), and 90% for sample stimuli within the green stimulus class. The difference in the mean percent of response to the two stimulus classes was 90%.

Intervention. In Figure 10(e), the far right graph of the first panel shows that Corbin needed 30 teaching opportunities to master the simple discrimination, involving the selection of "Green" choice at the presentations of the sample S_A (i.e., 499 nm).

Generalization testing. In Figure 10(e), the center graph of the first panel shows that subsequent to intervention, Corbin's performance during the implementation of generalization gradient assessment improved to be exactly identical to the ideal generalization gradient. The percent of the "Green" choice in the presence of the sample S_A (i.e., 499 nm) increased from 40% during baseline to 100%. The mean percent of the "Green" choice was 0% for sample stimuli associated with the blue stimulus class, which was the same as that during baseline (i.e., an indication of no over-generalization of the "Green" choice). The mean percent of the "Green" choice was 100% for sample stimuli associated with the green stimulus class, which was an increase from 90% during baseline. The difference in the mean percent of response to the two stimulus classes was the perfect 100%, which was an increase from 90% during baseline.

Maximal-Difference Conditional Discrimination Condition

The second panel of Figure 10(e) shows Corbin's performance on the maximal-difference conditional discrimination condition (assigned with the orange-to-red color set).

Baseline. In Figure 10(e), the far left graph of the second panel shows that during baseline Corbin's performance was not in the ideal sigmoidal form, although his responding to the two stimulus classes was visibly differentiated. The percent of the “Red” choice in the presence of the sample S_A (i.e., 631 nm) was 20% of opportunities. The mean percent of the “Red” choice was 10% for sample stimuli within the orange stimulus class (i.e., 619, 622, 625, 628 nm), and 70% for sample stimuli within the red stimulus class (i.e., 634, 637, 641, 645 nm). The difference between the mean percent of response to the two stimulus classes was 60%.

Intervention. In Figure 10(e), the far right graph of the second panel shows that Corbin required 170 teaching opportunities to master the maximal-difference conditional discrimination between the sample S_A of 631 nm (representing red) and the sample S_B of 619 nm (representing orange).

Generalization testing. In Figure 10(e), the center graph of the second panel shows that subsequent to intervention, a response bias toward the “Red” choice was observed, with an increase of the “Red” responses to both stimulus classes. The percent of the “Red” choice given the sample S_A of 631 nm increased from 20% during baseline to 100%. The mean percent of the “Red” choice for sample stimuli associated with the orange stimulus class increased from 10% during baseline to 45%, which was an indication of over-generalization of the “Red” choice. The mean percent of the “Red” choice for sample stimuli associated with the red stimulus class increased from 70% during baseline to 95%. The difference between the mean percent of response to the two stimulus classes was 50%, which was a decrease from 60% during baseline.

Minimal-Difference Conditional Discrimination Condition

The third panel of Figure 10(e) shows Corbin’s performance on the minimal-difference conditional discrimination condition (assigned with the yellow-to-orange color set).

Baseline. In Figure 10(e), the far left graph of the third panel shows that during baseline Corbin’s performance was not in the ideal sigmoidal form. Especially his performance within the orange stimulus class (i.e., 597, 600, 603, 606 nm) was variable and demonstrated a trend in a direction opposite to the ideal one. He selected the “Orange” choice in the presence of the sample

S_A (i.e., 594 nm) during 80% of opportunities. The mean percent of the "Orange" choice was 5% for sample stimuli within the yellow stimulus class (i.e., 580, 584, 588, 591 nm), and 60% for sample stimuli within the orange stimulus class. The difference in the mean percent of response to the two stimulus classes was 55%.

Intervention. In Figure 10(e), the far right graph of the third panel shows that Corbin required 90 teaching opportunities to master the minimal-difference conditional discrimination between the sample S_A of 594 nm (representing orange) and the sample S_B of 588 nm (representing yellow).

Generalization testing. In Figure 10(e), the center graph of the third panel shows that following mastery Corbin's performance during the implementation of generalization gradient assessment demonstrated both improvement and deterioration. In general a response bias toward the "Orange" choice was observed, with an increase of the "Orange" responses to both stimulus classes. He selected the "Orange" choice in the presence of the sample S_A during 40% of opportunities, which was a decrease from 80% during baseline. The mean percent of "Orange" choice for sample stimuli associated with the yellow stimulus class was 20%, which was an increase from 5% during baseline. This was an indication of over-generalization of the "Orange" choice. The mean percent of "Orange" choice for sample stimuli associated with the orange stimulus class was 85%, which was an increase from 60% during baseline. The difference in the mean percent of response to the two stimulus classes was 65%, which was an increase from 55% during baseline.

Comparison Across Experimental Conditions

Corbin required the least number of teaching opportunities (30) to master the simple discrimination with the minimal-difference conditional discrimination being mastered with 90 teaching opportunities. The maximal-difference conditional discrimination required the greatest number of teaching opportunities (170) to reach mastery. This is inconsistent with the predicted order given the available literature (Eckerman, 1970; Galizio, 1980; LaBerge, 1961).

From baseline to generalization testing, the simple discrimination was most effective in enhancing the overall performance toward the ideal performance, followed by the minimal-difference conditional discrimination. The maximal-difference conditional discrimination was least effective in enhancing the overall performance. The simple discrimination and maximal-difference conditional discrimination enhanced Corbin's correct responses to the corresponding sample S_A, whereas the minimal-difference conditional discrimination reduced Corbin's correct responses to the corresponding sample S_A. All three conditions enhanced the performance within corresponding stimulus Class A. The simple discrimination did not change the performance within stimulus Class B, and the two conditional discriminations (especially the maximal-difference one) resulted in some deterioration of the performance within corresponding stimulus Class B.

Participant Comparison

In this section, results are presented across all five participants with respect to the relative efficiency of discrimination learning, and the change of the generalization gradient performance between baseline and generalization testing. Finally, a brief summary is presented.

Teaching Opportunities to Mastery Criterion During Intervention

Table 18 shows the teaching opportunities to mastery criterion averaged across five participants. Overall, the five participants mastered the simple discrimination during an average of 74 teaching opportunities (range: 30-160 teaching opportunities). They mastered the maximal-difference conditional discrimination during an average of 230 teaching opportunities (range: 30-470 teaching opportunities). They mastered the minimal-difference conditional discrimination during an average of 176 teaching opportunities (range: 90-310 teaching opportunities).

On average, these five participants learned the simple discrimination most quickly, followed by the minimal-difference conditional discrimination, and then the maximal-difference conditional discrimination. This trend was applicable to three participants (i.e., Bryan, James, and Corbin), but not to the other two participants (i.e., Bruce and Gabi). For Bruce, he also learned the simple discrimination most quickly, but he learned the maximal-difference conditional discrimination faster than the minimal-difference one. For Gabi, she learned the maximal-difference conditional discrimination most quickly, followed by the simple discrimination, and then the minimal-difference conditional discrimination.

Change of Generalization Gradient Performance From Baseline to Generalization Testing

The change of generalization gradient performance from baseline to generalization testing was described across five participants from the following four aspects of performance that included, (a) the overall performance, (b) the responding to the sample S_A (i.e., the PSI stimulus), (c) performance within stimulus Class A, and (d) performance within stimulus Class B.

In terms of the overall performance, as Figure 8 shows, in the ideal sigmoidal form the percent of the target choice is 100% for the PSI stimulus and sample stimuli within the stimulus Class A, and the percent of the target choice is 0% for sample stimuli within the stimulus Class B. From the perspective of stimulus control, the closer an actual generalization gradient gets to the ideal one, the larger degree of stimulus control it approximates and the better the overall performance it represents. As Figures 10(a) to 10(e) show, for Bruce the generalization gradients subsequent to the simple discrimination and the minimal-difference conditional discrimination resulted in a larger degree of stimulus control compared to the baseline performance, with both more desired responding within each of two stimulus classes and better discriminated responses between two stimulus classes. Instead, the maximal-difference conditional discrimination resulted in over-use of the "Orange" choice. For Gabi, the generalization gradients subsequent to the two conditional discriminations resulted in a larger degree of stimulus control compared to the baseline performance, whereas the simple discrimination resulted in over-use of the “Red” choice. For Bryan, the generalization gradients subsequent to each of the three interventions demonstrated a greater degree of stimulus control compared to the baseline performance. For James, none of the generalization gradients subsequent to the three interventions demonstrated enhanced stimulus control compared to the baseline performance. For Corbin, the generalization gradient following the simple discrimination resulted in enhanced stimulus control compared to the baseline performance, whereas both the maximal- and minimal-difference conditional discriminations resulted in over-use of one response choice. Overall, apart from James, four of the remaining participants improved their overall performance toward the ideal sigmoidal form (i.e., achieved an enhanced stimulus control) during generalization testing on at least one discrimination intervention condition.

With respect to the correct responses given the sample S_A (i.e., the PSI stimulus of each color set), results were somewhat mixed across participants. For Bruce, Gabi, and James, the simple discrimination was most effective in enhancing the correct responses in the presence of the sample S_A. For Corbin, both the simple discrimination and the maximal-difference conditional discrimination were more effective in enhancing the correct responses in the presence of the sample S_A. For Bryan, the two conditional discriminations were more effective in enhancing the correct responses in the presence of the sample S_A. In addition, the percent of correct responses given the sample S_A decreased from baseline to generalization testing on the minimal-difference conditional discrimination condition for all five participants except Bryan.

Some consistent results were observed with regard to the performance within stimulus Class A across five participants. The simple discrimination enhanced the performance within stimulus Class A for all five participants. Both the maximal-difference and the minimal-difference conditional discriminations were effective in enhancing the performance within stimulus Class A for all participants except James.

Less consistent results were observed in terms of the performance within stimulus Class B across five participants. For Bruce, the simple discrimination and the minimal-difference conditional discrimination were more effective in enhancing the performance within stimulus Class B. For Gabi and Bryan, the two conditional discriminations (especially the minimal-difference one) were more effective in enhancing the performance within stimulus Class B. For James, none of the three conditions were effective in enhancing the performance within stimulus Class B. Corbin had an ideal or close to ideal performance within stimulus Class B during baseline across three conditions, and a deterioration of the performance within stimulus Class B was observed on the two conditional discrimination conditions, especially on the maximal-difference conditional discrimination condition.

Summary

The simple discrimination condition required the fewest teaching opportunities to mastery criterion. Subsequent to intervention, all three conditions could enhance participants' overall performance but their relative effectiveness varied across participants. In particular, the simple discrimination enhanced the overall performance toward the ideal generalization gradient for three participants (i.e., Bruce, Bryan, and Corbin). In particular, Corbin's performance on the post acquisition generalization gradient was exactly the same as the ideal one, reflecting perfect stimulus discrimination and stimulus generalization across the assigned blue-to-green color continuum. The maximal-difference conditional discrimination enhanced the overall performance toward the ideal generalization gradient for two participants (i.e., Gabi and Bryan). The minimal-difference conditional discrimination enhanced the overall performance toward the ideal generalization gradient across three participants (i.e., Bruce, Gabi, and Bryan).

More specifically, the simple discrimination was more effective in improving in the correct responding given the sample S_A (i.e., the most ambiguous stimulus) than the two conditional discriminations, whereas the minimal-difference conditional discrimination tended to result in a decrease in the correct responding given the sample S_A. The three conditions were generally equivalent in improving the performance within stimulus Class A. The minimal-difference conditional discrimination was relatively more effective in improving the performance within stimulus Class B than the other two conditions.

Chapter V: DISCUSSION

Findings relevant to the research questions of the current study are first discussed, followed by a summary of the limitations of the current study. Finally, a discussion of the major implications for research and practice is provided.

Efficiency of the Acquisition of Discrimination Learning

The first research question was "Does a maximal-difference, or a minimal-difference (between two teaching stimuli S_A and S_B) conditional discrimination condition, or a simple discrimination condition result in the fewest teaching opportunities to reach mastery criterion?" The original hypothesis was that the simple discrimination would require of the fewest teaching opportunities to reach mastery criterion, followed by a maximal-difference conditional discrimination condition, and then a minimal-difference conditional discrimination condition. The hypothesis was partially confirmed. Results indicated that in general the simple discrimination condition required a smaller number of teaching opportunities to criterion during acquisition intervention than the two conditional discrimination conditions. This finding is consistent with the results reported by Galizio (1980). This outcome is consistent not only with existing evidence but also understandable given that simple discriminations have been reported to be cognitively less demanding than conditional discriminations (Axe, 2008; Reichle & Wilkinson, 2012).

Less conclusive data were obtained with regard to the relative efficiency between maximal- and minimal-difference in teaching stimuli compared in the two conditional discrimination conditions that were implemented. Two participants mastered the maximal-difference conditional discrimination faster than the minimal-difference one. Again, this outcome is consistent with extant literature suggesting that greater differences among teaching stimuli speed acquisition (Eckerman, 1970; LaBerge, 1961). However, somewhat more difficult to explain was the outcome that three participants mastered the minimal-difference condition with fewer teaching opportunities than the maximal-difference condition. This latter finding requires more careful consideration.

It is possible that the differences between stimuli in maximal and minimal conditional discrimination conditions might have been relatively equivalent in their difficulty. Although the interval between two teaching stimuli varied in the maximal and minimal conditions; for some learners, their difference may not have been sufficiently large to exert a differential influence on the learning outcomes between the two conditions. Horner et al. (1986) also reported equivalent acquisition performance between a maximal-difference and a minimal-difference conditional discrimination conditions, although the minimal-difference conditional discrimination actually involved a greater number of teaching examples, which was purported by the authors to be theoretically more difficult.

In summary, the simple discrimination condition was most quickly acquired. The two conditional discriminations produced mixed results in terms of the speed with which they were acquired. Greater contrast between teaching stimuli for each of the two conditional discrimination conditions might have created greater differentiation in performance between the two conditions.

Effects of the Difference Between Teaching Stimuli in Conditional Discrimination Interventions on the Generalization Performance Subsequent to Acquisition

The second research question was “Does a maximal-difference, or a minimal-difference (between two teaching stimuli S_A and S_B) conditional discrimination condition, or a simple discrimination condition best enhance the generalization performance?” The original hypothesis was that the two conditional discrimination conditions would better enhance generalization performance than the simple discrimination condition. The minimal-difference conditional discrimination condition would better enhance the generalization performance than the maximal-difference conditional discrimination condition. The hypothesis was also partially confirmed.

Comparing Conditional Discrimination Conditions with Simple Discrimination Condition

Results from the current study did not provide evidence that a conditional discrimination condition resulted in better generalization performance than a simple discrimination condition. Instead, results suggested that the simple discrimination condition enhanced the generalization performance toward the ideal generalization gradient (or resulted in an enhanced stimulus control) for three participants (i.e., Bruce, Bryan, and Corbin), which was at least as effective as the two conditional discrimination conditions. To be specific, the simple discrimination condition was at least as effective as the other two intervention conditions in achieving more desired generalization over the PSI stimulus and those stimuli within stimulus Class A (i.e., the stimuli categorized as sharing the color category assigned to the PSI stimulus for teaching purposes). Moreover, participant performance in the simple discrimination condition was reasonably similar to that in the maximal-difference conditional discrimination condition in terms of enhancing performance within stimulus Class B (i.e., the stimuli categorized as not sharing the color category assigned to the PSI stimulus for teaching purposes), although the simple discrimination was less effective than the minimal-difference conditional discrimination in enhancing performance within stimulus Class B.

Two major learning mechanisms help explain the preceding outcome. One is related to the adaptation level theory (Helson, 1964; Thomas et al., 1991), and the other is related to the method used in implementing the simple discrimination condition.

According to the adaptation level theory (Helson, 1964; Thomas et al., 1991), for the simple discrimination condition, the single teaching sample S_A (i.e., the PSI stimulus) was the adaptation level during both intervention and generalization testing. Thus, participants did not need to develop a new adaptation level during generalization testing. They could simply respond using a rule indicating that, "If the stimulus presented is the sample S_A or more similar to the sample S_A, respond to one switch choice; If the stimulus presented is less similar to the sample S_A, respond to the other switch choice." This may have reduced participant's cognitive load.

For example, during an intervention a participant might be taught to emit response A when he or she lifts an item that is 11 pounds or above, and emit response B when he or she lifts an item that is lighter than 11 pounds. If the stimuli utilized during generalization testing were 10 items of different weights, including 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 pounds. The adaptation level during this generalization testing is 11 pounds, identical to the teaching stimulus value during intervention. Alternatively, if the stimuli utilized during generalization testing were 10 items of different weights, including 1, 4, 8, 12, 16, 20, 24, 28, 32, and 36 pounds. The adaptation level during this generalization testing is 18.1 pounds, higher than the teaching stimulus value of 11 pounds during intervention. It is anticipated that the participant would demonstrate more accurate performance in the former generalization testing and have more difficulty in the latter generalization testing due to the potential confusion derived from the formation of a new adaptation level during generalization testing.

Second, the finding that the simple discrimination was able to enhance generalization performance during generalization testing in this study may support the proposition that the expression of stimulus control in multiple-response situations differs from that in the single-response case (Cross & Lane, 1962). That is, the simple discrimination task in the current investigation was somewhat different from the simple discrimination task in the study by Galizio (1980), although in both tasks the positive (S+) and negative (S-) functions of stimuli remained the same from opportunity to opportunity (Catania, 1998). In Galizio's study the simple discrimination involved two teaching stimuli but only one response choice (e.g., pressing a key when a tone of 1200 Hz was presented but not pressing the key when a tone of 1400 Hz was presented), and the resulting generalization gradient was in a peaked shape. In this investigation the simple discrimination involved only one teaching stimulus but two response choices. In the study by Cross and Lane (1962), they pointed out that generalization gradients obtained following a multiple-response discrimination learning divided the stimulus continuum into response-specific stimulus classes, which contrasted with the peaked- shape generalization gradients obtained subsequent to a single-response discrimination learning. Nonetheless, in their study the multiple-response discrimination learning was actually a two-sample two-choice conditional discrimination. Therefore, whether the adoption of multiple responses itself had played a critical role in the enhanced generalization gradient performance following a one-sample two-choice simple discrimination in this study remains unclear.

To some extent, the finding that the simple discrimination in this study also improved participants' generalization performance toward the ideal generalization gradient is somewhat at odds with the existing literature on general case instruction that emphasizes the selection and implementation of both positive and negative teaching exemplars to enhance the generalization performance (Chadsey-Rusch & Halle, 1992; Horner et al., 1982; O'Neill & Reichle, 1993). However, it is unclear which critical component(s) were responsible for the outcomes obtained. The potential critical components may include (a) the use of the PSI stimulus as the only teaching stimulus that was also the adaptation level during generalization testing, (b) the adoption of two response choices itself, and (c) the relatively simplified teaching stimuli (i.e., color stimuli that varied on one major dimension of the wavelength). Without further research, caution should be taken prior to making a conclusion that a simple discrimination involving a singular teaching stimulus was as efficient as a conditional discrimination in leading to a balanced stimulus discrimination and stimulus generalization, or well-separated stimulus classes.

Comparing Maximal-Difference with Minimal-Difference Conditional Discriminations

The current study found that the maximal- and minimal-difference conditional discriminations were generally equivalent in enhancing the overall performance (or the degree of stimulus control), although the minimal-difference condition demonstrated slight superiority.

Generally speaking, the minimal-difference conditional discrimination led to a decrease in the correct responding given the sample S_A while the maximal-difference conditional discrimination led to an increase in the correct responding given the sample S_A. The two conditional discriminations were generally equivalent in enhancing the performance within stimulus Class A. However, the minimal-difference conditional discrimination appeared to be relatively more effective in enhancing performance within stimulus Class B than the maximal-difference conditional discrimination. This is consistent with the findings by Horner et al. (1986) who found that both maximal-difference and minimal-difference conditional discrimination interventions enhanced the generalization performance of learners with IDD within one stimulus class (i.e., successfully accepting the untaught items that matched a picture card in an untaught store); nonetheless, a minimal-difference conditional discrimination was required for these learners to also enhance their generalization performance within the other stimulus class (i.e., correctly rejecting the untaught items that did not match a picture card in an untaught store).

In part, the findings just described could account for the peak shift phenomenon frequently noted in basic research literature involving simple discrimination (e.g., Akins et al., 1981; Cheng et al., 1997; Dickinson & Hedges, 1986; Dougherty & Lewis, 1991; Marsh, 1972; Wheatley & Thomas, 1974; Wilkie, 1972; Wills & Mackintosh, 1998). According to the peak shift phenomenon, in a simple discrimination as S+ and S- get increasingly similar during acquisition intervention, generalization outcomes result in less over-generalization to stimuli less similar to S+ as well as less under-generalization to stimuli more similar to S+ (Ghirlanda & Enquist, 2003). A similar finding regarding peak shift subsequent to a conditional discrimination intervention was reported by Galizio (1980). Although the conditional discrimination procedures implemented in this study were different from those implemented by Galizio (1980), the finding from the current study suggests that as the difference between two teaching stimuli decreased, at least the performance for one stimulus class was relatively more enhanced with improved generalization.

Additionally, the current study did not provide sufficient evidence to support the proposition that as the difference between two teaching stimuli decreased the performance within stimulus Class A was enhanced. Specifically, performance within stimulus Class A in both the minimal- and maximal-difference conditional discrimination conditions was equivalently enhanced. Moreover, the minimal-difference condition resulted in a decrease in the correct responding given the sample S_A while the maximal-difference one led to an increase in the correct responding given the sample S_A. The correct response to the sample S_A was identical to the correct response to the sample stimuli from stimulus Class A. Therefore, if the performance given the sample S_A were analyzed as part of the performance within stimulus Class A, there was less sufficient evidence to suggest that the minimal-difference conditional discrimination resulted in better performance within stimulus Class A than the maximal-difference one. With that being said, caution should be taken in concluding that the maximal-difference conditional discrimination was more effective in enhancing the performance within stimulus Class A than the minimal-difference condition. Baseline responding to the sample S_A in the minimal-difference condition was better than it was in the other two conditions for all five participants (especially for Bryan and Corbin). Thus, the change of the correct responding given the sample S_A subsequent to the intervention might have been influenced by the differential baseline responding across conditions.

In summary, the current study demonstrated that a conditional discrimination condition did not result in better generalization performance than a simple discrimination condition. As the difference between the teaching stimuli representing stimulus Class A and stimulus Class B decreases, generalization performance within at least one stimulus class was enhanced.

Limitations

Several procedural features of the current investigation may pose threats to the internal validity of the current study. As already mentioned, baseline performance given the sample S_A (i.e., the PSI stimulus in the study) was not quite equivalent across three intervention conditions (especially for the minimal-difference conditional discrimination condition). This could have influenced generalization gradient performance during generalization testing. Second, the current study compared the effects of three experimental conditions, but did not involve a control condition (i.e., no intervention condition). The lack of a true control condition may have restricted the detection of potential confounding variables such as maturity, interaction between conditions, and any other possible learning that was not a function of instruction. Although all sessions were conducted in the participants' home, the likelihood of learners being inadvertently exposed to experimental stimuli was relatively minimal in this study with the way the experimental stimuli were presented taken into account. The counterbalancing presentations of intervention conditions could help mitigate the concerns around potential interaction between conditions.

The reinforcement schedule used in this study was individualized to meet each learner's needs based on his or her history in prior learning tasks. Tangible reinforcement was applied to two participants (i.e., Bryan and James), whereas a combination of conditioned reinforcement and tangible reinforcement was applied to the remaining participants (i.e., Bruce, Gabi, and Corbin). Given the within subject design this could not have been a threat to internal validity although it is possible that it could have had some influence upon the comparisons in performance across learners.

In addition to the limitations surrounding internal validity, there may have been several limitations related to external validity. This study involved only five young learners with moderate to severe I/DD who were all diagnosed with Down syndrome. This limits the generalizability of the findings to other populations. Second, the color stimuli used as the experimental materials differed in only one major stimulus dimension (i.e., wavelength adjusted with luminosity). Thus, the results may not be generalizable to tasks involving different or more complicated stimuli. Finally, the current study involved a basic two-choice matching task that was implemented in a simultaneous and non-successive matching format. Therefore, the findings may not be readily generalized to other types of conditional discrimination tasks such as those involving more than two samples and/or more than two choices, those implemented in a delayed matching format (i.e., involving a delay between the presentation of the sample and the choice), and those implemented in a successive matching format (i.e., involving non-concurrent presentations of choices).

Implications for Research

Given the relatively heterogeneous results reported, future research should consider replications with the addition of a true control task. Additionally better equating the baseline performance across conditions (especially the baseline performance in the presence of the PSI stimulus as the teaching stimulus) may also be warranted (although difficult to execute). The potential factors (e.g., the complexity of experimental stimuli, the use of the PSI stimulus as the teaching stimulus, or the number of choices) accounting for the enhanced stimulus control subsequent to a simple discrimination could be further examined. Additionally, research could be implemented to explore whether incorporating the PSI stimulus as the teaching stimulus in a conditional discrimination intervention would be conducive to the generalization outcomes. Furthermore, future researchers may wish to consider response latency as an additional dependent measure, given its potential value in the detection of more subtle changes in performance (Capehart & Peace, 1968). In addition, future research could explore approaches to addressing less desired generalization errors or failures subsequent to initial discrimination learning (e.g., James' performance in this study), which was not the focus of the current study.

Second, future research could examine whether similar outcomes would be obtained when the conditional discrimination task involved other types of experimental stimuli. In particular, the current study suggests that as the difference between teaching stimuli in a conditional color discrimination task decreased, generalization performance was enhanced in at least one stimulus class. This finding is consistent with the extant small body of applied behavioral literature (Horner et al., 1986). Researchers could utilize different and possibly more complex stimuli to examine this phenomenon. These stimuli could be those varying on more than one stimulus dimension (e.g., varying on both visual and auditory dimensions, or varying on visual, auditory, and tactile dimensions) and those frequently encountered in real life settings (e.g., reading materials, word problems, and social contexts for communication). More evidence accumulated in this area could help enhance the generality of this finding obtained in this study and expand the literature on how differences of the sampled teaching stimuli may influence the generalization performance.

Third, future researchers could expand investigations on how other intervention variables involved in conditional discrimination tasks may impact the discrimination and generalization performance using a research design the same as or similar to the one used in the current study and employing a generalization gradient as a investigation tool. In relation to the well-documented general case instruction (O'Neill & Reichle, 1993), these intervention variables may include but not limited to, the number of teaching stimuli, the teaching sequence of teaching stimuli, the reinforcement schedule, and the intervention intensity. For instance, the existing basic behavioral evidence indicates that different reinforcement schedules may exert influence upon the generalization gradient performance (Guttman, 1959; Hearst, Koresko, & Poppen, 1964; Yarczower, Dickson, & Gollub, 1966; Zeiler, 1969). It is important to examine whether similar findings would be obtained in applied behavioral research for the expansion of empirical evidence in this field. In a similar vein, future researchers could explore the impacts of other generalization-enhancing strategies on the generalized outcomes using the same experimental procedures as those in the current study. For example, future researchers could examine how specific tactics (e.g., exploiting current functional contingencies, incorporating functional mediators such as rules and language) advocated by Stokes and Osnes (1989) may impact the generalized outcomes.

Fourth, although rarely addressed in the existing literature, future researchers may wish to further investigate options for measuring generalization. As this study demonstrates, compared to the conventionally dichotomous documentation of generalization performance in applied behavioral research, the use of a generalization gradient allows for more sensitive and representative capture of the generalized and discriminated outcomes across a specific continuum of probes (Honig & Urcuioli, 1981). However, some limits of generalization gradients may exist. Specifically, in using a generalization gradient the stimulus dimension(s) need to be carefully characterized so as to be amenable to the alignment on a scale, with the interval between two points along the scale clearly defined. Additionally, evidence addressing approaches to visually and quantitatively analyzing the generalization gradient performance remains sparse.

Finally, future researchers should consider investigations of generalization in its own right, especially in the field of applied behavioral research. The significant role of generalization in survival and evolution has been much documented (Pavlov, 1927; Shepard, 1987; Vervliet et al., 2011; Wisniewski et al., 2009), however, the majority of the extant work investigating how different parameters may influence generalization performance has been grounded in the basic behavioral research (see reviews by Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981). Much more work in the applied arena is needed. In this way, evidence on generalization patterns and generalization error patterns may be gradually accumulated. In addition, although the current study took the baseline generalization performance into consideration, it did not investigate the maintenance of generalization performance. Given the theoretical and practical significance of maintained generalization (Hinson & Malone, 1980; Howard, 1979; Malone & Staddon, 1973; Schuster & Gross, 1969; Stokes & Baer, 1977; Stokes & Osnes, 1989; Tennison & Hinson, 1993; White & Thomas, 1979; Winograd, 1965), future researchers should also make endeavors to examine the temporal dynamics of generalization.

Implications for Practice

Interventionists should not overlook the importance of carefully selecting the positive and negative teaching examples in teaching a wide range of conditional discrimination skills for well-generalized performance. As the results from this study indicate, as the positive and negative teaching examples get more similar, a learner with IDD may be able to further improve his or her generalization performance within at least one stimulus class. The consideration of more carefully selecting teaching stimuli with greater similarity probably becomes more important if the initial conditional discrimination is insufficient in establishing acceptable discrimination and generalization performance. In particular, interventionists may consider incorporating the stimulus or stimuli that present to be quite challenging for the learners to make discriminations (e.g., the PSI stimulus in the current study) to further enhance the stimulus discrimination and generalization outcomes.

It would be also important to develop effective and "user-friendly" frameworks or processes for practitioner use in implementing general case instruction (O'Neill, Faulkner, & Horner, 2000). In particular, in the current study all teaching and testing sessions were completed through the laptop presentations, the sessions were more of structured in nature, although still in the participants' naturalistic homes. It would be helpful for practitioners to look at ways to combine more structured training sessions into the naturally occurring activities across a learner's days (e.g., the combination of structured training and naturalistic approaches; Kaiser & Roberts, 2013; O'Neill et al., 2000). In this manner, teaching can become more manageable and the practitioners would be less reluctant to implement it.

Conclusion

This study was the first investigation to specifically examine the generalization performance in a conditional discrimination task via a generalization gradient for learners with moderate to severe IDD through an adapted alternating treatment single case experimental design by manipulating one specific variable, that is, the difference between the teaching stimuli involved in the conditional discrimination task. The current study provides some preliminary evidence on how differences between teaching stimuli in a conditional discrimination task influenced the discrimination learning and generalization performance for learners with IDD. In general, these five participants required the least amount of teaching opportunities to master the simple discrimination. Four out of the five participants demonstrated improvement in their overall generalization performance after intervention on at least one intervention condition. However, there was not sufficient evidence to support that a simple discrimination condition was less effective in enhancing the overall generalization performance subsequent to the intervention than the two conditional discrimination conditions. As the difference between two teaching stimuli in the conditional discrimination task decreased, an enhanced generalization performance was observed in one stimulus class. How these findings could be extended for greater internal validity and external validity is an important project in future.

References

Ahearn, W. H., MacDonald, R. P. F., Graff, R. B., & Dube, W. V. (2007). Behavior analytic teaching procedures: Basic principles, empirically derived practices. In P. Sturmey & A. Fitzer (Eds.). *Autism Spectrum Disorders: Applied Behavior Analysis, Evidence, and Practice*. Austin, TX: Pro-Ed.

Albin, R. W., & Horner, R. H. (1988). Generalization with precision. In Horner, R. H., Dunlap, G., & Koegel, R. L. (eds.), *Generalization and maintenance: Life-style changes in applied settings*, Baltimore, Paul H. Brookes, pp. 99-120.

Alloway, T. P. (2010). Working memory and executive function profiles of students with borderline intellectual functioning. *Journal of Intellectual Disability Research*, *54*, 448-456. doi: 10.1111/j.1365-2788.2010.01281.x

Andrews, G., Halford, G. S., & Boyce, J. (2012). Conditional discrimination in young children: The roles of associative and relational processing. *Journal of Experimental Child Psychology*, *112*(1), 84-101. doi:10.1016/j.jecp.2011.12.004

Arkins, F. R., Drew Gouvier, W., & Lyons, J. E. (1981). The effects of S+ and S- alternation on stimulus control and the peak shift. *Journal of General Psychology, 105*, 101-114. doi:10.1080/00221309.1981.9921061

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis*, *12*, 505-522.

Axe, J. B. (2008). Conditional discrimination in the intraverbal relation: A review and recommendations for future research. *The Analysis of Verbal Behavior, 24*, 159-174.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91-97. doi:10.1901/jaba.1968.1-91

Bailey, S. L. (1981). Stimulus overselectivity in learning disabled children. *Journal of Applied Behavior Analysis, 14*(3), 239-248. doi:10.1901/jaba.1981.14-239

Baron, A. (1973). Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology, 101*, 337-342. doi:10.1037/h0035206

Belmont, J. M., & Butterfield, E. C. (1969). The relations of short-term memory to development and intelligence. In L. P. Litsitt & H. W. Reese (Eds.), *Advances in child development and behavior* (Vol. 4). New York, NY: Academic Press.

Bizo, L. A., & McMahon, C. V. (2007). Temporal generalization and peak shift in humans. *Learning & Behavior, 35*(2), 123-130.

Bloomfield, T. M. (1967). A peak shift on a line-tilt continuum. *Journal of the Experimental Analysis of Behavior, 10*, 361-366. doi:10.1901/jeab.1967.10-361

Blough, D. S. (1973). Two-way generalization peak shift after two-key training in the pigeon. *Animal Learning & Behavior*, *1*(3), 171-174.

Brennan, J. F., & Riccio, D. C. (1972). Stimulus control of shuttle avoidance in young and adult rats. *Canadian Journal of Psychology, 26*, 361-373. doi:10.1037/h0082443

Browder, D. M., & Grasso, E. (1999). Teaching money skills to individuals with mental retardation: A research review with practical applications. *Remedial and Special Education, 20*(5), 297-308. doi:10.1177/074193259902000506

Brownell, R. (2000). *Expressive one-word picture vocabulary test* (3rd ed.). Novato, Calif: Academic Therapy Publications.

Brownell, R. (2000). *Receptive one-word picture vocabulary test* (3rd ed.). Novato, Calif: Academic Therapy Publications.

Capehart, J., & Pease, V. (1968). An application of adaptation-level theory to transposition responses in a conditional discrimination. *Psychonomic Science*, *10*(4), 147-148. doi: 10.3758/bf03331453

Carlesimo, G. A., Marotta, L., & Vicari, S. (1997). Long-term memory in mental retardation: Evidence for a specific impairment in subjects with Down's syndrome. *Neuropsychologia, 35*, 71-79.

Carter, E. W., Sisco, L. G., Chung, Y-C., & Stanton-Chapman, T. L. (2010). Peer interactions of students with intellectual disabilities and/or autism: A map of the intervention literature. *Research & Practice for Persons with Severe Disabilities, 35*(3-4), 63-79. doi:10.2511/rpsd.35.3-4.63

Carter, D. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior*, *29*(3), 565-601. doi:10.1901/jeab.1978.29-565

Catania, A. C. (1998). *Learning* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Catania, A. C. (1992). *Learning* (3rd ed.). Englewood Cliffs, NJ: Prentice—Hall.

Cederborg, A. C., La Rooy, D., & Lamb, M. E. (2008). Repeated interviews with children who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *21*, 103-113. doi:10.1111/j.1468-3148.2007.00372.x

Cha, K-H., & Merrill, E. C. (1994). Facilitation and inhibition effects in visual selective attention processes of persons with and without mental retardation. *American Journal on Mental Retardation*, *98*, 594-600.

Chadsey-Rusch, J., Drasgow, E., Reinoehl, B., Halle, J., & Collet-Klingenberg, L. (1993). Using general-case instruction to teach spontaneous and generalized requests for assistance to learners with severe disabilities. Research and Practice for Persons with Severe Disabilities, 18(3), 177-187. doi:10.1177/154079699301800304

Chadsey-Rusch, J., & Halle, J. (1992). The application of general-case instruction to the requesting repertoires of learners with severe disabilities. *Journal of the Association for Persons with Severe Handicaps, 17*(3), 121-132. doi:10.1177/154079699201700301

Cheng, K., Spetch, M. L., & Johnston, M. (1997). Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology, 23*(4), 469-481.

Chezan, L. C., Drasgow, E., & Marshall, K. J. (2012). A report on using general-case programming to teach collateral academic skills to a student in a postsecondary setting. *Focus on Autism and Other Developmental Disabilities, 27*(1), 22-30. doi:10.1177/1088357611428334

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

Cross, D. V., & Lane, H. L. (1962). On the discriminative control of concurrent responses: The relations among response frequency, latency, and topography in auditory generalization. *Journal of Experimental Analysis of Behavior*, *5*, 487-496. doi:10.1901/jeab.1962.5-487

Cumming, W. W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior*, *4*(3), 281-284. doi:10.1901/jeab.1961.4-281

Day, H. M., & Horner, R. H. (1986). Response variation and the generalization of a dressing skill: Comparison of single instance and general case instruction. *Applied Research in Mental Retardation*, *7*, 189-202. doi:10.1016/0270-3092(86)90005-6

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*(4), 519-533. doi:10.1901/jaba.1996.29-519

Derenne, A., Breitstein, R. M., & Cicha, R. J. (2008). Shifts in postdiscrimination gradients within a stimulus dimension based on female waist-to-hip ratios. *The Psychological Record, 58*, 51-60.

Derenne, A. (2010). Shifts in postdiscrimination gradients within a stimulus dimension based on bilateral facial symmetry. *Journal of the Experimental Analysis of Behavior, 93*(3), 485-494. doi:10.1901/jeab.2010.93-485

de Souza, S., Goyos, C., Silvares, E. F. M., & Saunders, R. R. (2007). Emergence of printing and spelling skills from constructed-response matching-to-sample instruction (CRMTS). *European Journal of Behavior Analysis*, *8*(1), 49-64. doi:10.1080/15021149.2007.11434273

Dickinson, J., & Hedges, D. G. (1986). Adaptation level as an explanation of the peak shift in generalization with movement stimuli. *Journal of Mathematical Biology, 18*, 101-110. doi:10.1080/00222895.1986.10735372

Dickson, C. A., Deutsch, C. K., Wang, S. S., & Dube, W. V. (2006). Matching-to-sample assessment of stimulus overselectivity in students with intellectual disabilities. *American Journal on Mental Retardation, 111*, 447-453. doi:10.1352/0895-8017(2006)111[447:MAOSOI]2.0.CO;2

Dickson, C. A., Wang, S. S., Lombard, K. M., & Dube, W. V. (2006). Overselective stimulus control in residential school students with intellectual disabilities. *Research and Intervention in Developmental Disabilities, 27*, 618-631. doi:10.1016/j.ridd.2005.07.004

Didden, R., Korzilius, H., van Oorsouw, W., & Sturmey, P. (2006). Behavioral treatment of challenging behaviors in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation, 111*(4), 290-298. doi:10.1352/0895-8017(2006)111[290:btocbi]2.0.co;2

Dougherty, D. M., & Lewis, P. (1991). Stimulus generalization, discrimination learning, and peak shift in horses. *Journal of the Experimental Analysis of Behavior, 56*, 97-104. doi:10.1901/jeab.1991.56-97

Doughty, A. H., & Hopkins, M. N. (2011). Reducing stimulus overselectivity through an increased observing-response requirement. *Journal of Applied Behavior Analysis*, *44*(3), 653-657. doi:10.1901/jaba.2011.44-653

Doughty, A. H., & Saunders, K. J. (2009). Decreasing errors in reading-related matching to sample using a delayed-sample procedure. *Journal of Applied Behavior Analysis*, *42*(3), 717-721. doi:10.1901/jaba.2009.42-717

Dube, W. V., Iennaco, F. M., & McIlvane, W. J. (1993). Generalized identity matching to sample of two-dimensional forms in individuals with intellectual disabilities. *Research in Developmental Disabilities*, *14*, 457-477. doi:10.1016/0891-4222(93)90038-l

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17*, 8-14.

Dube, W. V., McDonald, S. J., McIlvane, W. J., & Mackay, H. A. (1991). Constructed-response matching to sample and spelling instruction. *Journal of Applied Behavior Analysis*, *24*(2), 305-317. doi:10.1901/jaba.1991.24-305

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 68*(3), 303-316. doi:10.1901/jeab.1997.68-303

Dube, W. V., & McIlvane, W. J. (2002). Reinforcer rate and stimulus control in discrimination reversal learning. *The Psychological Record*, *52*, 405-416.

Dube, W. V., & Serna, R. W. (1998). Re-evaluation of a programmed method to teach generalized identity matching to sample. *Research in Developmental Disabilities*, *19*, 347-379.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). Transformation of avoidance response functions in accordance with the relational frames of Same and

Opposite. *Journal of the Experimental Analysis of Behavior*, *88*, 249-262. doi:10.1901/jeab.2007.22-07

Duker, P., Didden, R., & Sigafoos, J. (2004). *One to one training: Instructional procedures for learners with developmental disabilities*. Austin, TX: Pro-Ed.

Eckerman, D. A. (1970). Generalization and response mediation of a conditional discrimination. *Journal of the Experimental Analysis of Behavior, 13*, 301-316. doi:10.1901/jeab.1970.13-301

Enabling Devices (2011). *Talking Icon*®. http://enablingdevices.com/catalog/assistive_technology_devices_used_in_education/multi-coms/talking-icon

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York, NY: Irvington Publishers.

Falcomata, T. S., & Wacker, D. P. (2013). On the use of strategies for programming generalization during functional communication training: A review of the literature. *Journal of Developmental and Physical Disabilities, 25*, 5-15. doi:10.1007/s10882-012-9311-3

Ferrari, C., de Rose, J., & McIlvane, W. J. (2008). A comparison of exclusion and trial-and-error procedures: Primary and secondary effects. *Exp Anal Hum Behav Bull*, *29*, 9-16.

Fields, L., Garruto, M., & Watanabe, M. (2010). Varieties of stimulus control in a matching-to-sample: A kernel analysis. *The Psychological Record*, *60*, 3-26.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior, 68*, 67-91. doi:10.1901/jeab.1997.68-67

Frank, A. J., & Wasserman, E. A. (2005). Associative symmetry in the pigeon after successive matching to sample training. *Journal of Experimental Analysis of Behavior, 84*(2), 147-165. doi:10.1901/jeab.2005.115-04

Freeman, F. G., Kramarcy, N. R., & Lee, J. (1973). Discrimination learning and stimulus generalization in rats with hippocampal lesions. *Physiology & Behavior, 11*(2), 273-275. doi: 10.1016/0031-9384(73)90362-4

Galizio, M. (1980). Conditional gradient displacements: The effects of conditional discrimination training on human auditory frequency generalization. *Journal of Experimental Psychology: Human Learning and Memory, 6*(6), 766-777. doi: 10.1037/0278-7393.6.6.766

Galizio, M. (1985). Human peak shift: Analysis of the effects of three-stimulus discrimination training. *Learning and Motivation*, *16*, 478-494. doi:10.1016/0023-9690(85)90028-1

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behavior, 66*, 15-36. doi:10.1006/anbe.2003.2174

Giangreco, M. F. (1983). Teaching basic photography skills to a severely handicapped young adult using simulated materials. *Research and Practice for Persons with Severe Disabilities*, *8*(1), 43-49. doi:10.1177/154079698300800106

Grant, D. S., & Spetch, M. L. (1991). Pigeons' memory for event duration: Differences between choice and successive matching tasks. *Learning and Motivation, 22*, 180-199. doi:10.1016/0023-9690(91)90022-z

Grusec, T. (1968). The peak shift in stimulus generalization: Equivalent effects of errors and noncontingent shock. *Journal of the Experimental Analysis of Behavior, 11*, 239-249. doi:10.1901/jeab.1968.11-239

Guttman, N. (1959). Generalization gradients around stimuli associated with different reinforcement schedules. *Journal of Experimental Psychology*, 58(5), 335-340. doi:10.1037/h0045679

Guttman, N., & Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*, 79-88. doi:10.1037/h0046219

Hahs, A. D. (2015). Teaching prerequisite perspective-taking skills to children with autism. *International Journal of Psychology and Behavioral Sciences, 5*(3), 115-120. doi:10.5923/j.ijpbs.20150503.02

Hammond, J. L., Hirt, M., & Hall, S. S. (2012). Effects of computerized match-to-sample training on emergent fraction-decimal relations in individuals with fragile X syndrome. *Research in Developmental Disabilities*, *33*(1), 1-11. doi:10.1016/j.ridd.2011.08.021

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology, 58*(5), 321-334. doi:10.1037/h0042606

Hardin, C. L., & Maffi, L. (1997). *Color categories in thought and language*. Cambridge, UK: Cambridge University Press.

Hearst, E. (1968). Discrimination training as the summation of excitation and inhibition. *Science, 162*, 1303-1306. doi:10.1126/science.162.3859.1303

Hearst, E., Koresko, M. B., & Poppen, R. (1964). Stimulus generalization and the response-reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 7(5), 369-380. doi:10.1901/jeab.1964.7-369

Heinemann, E. G., & Chase, S. (1970). On the form of stimulus generalization curves for auditory intensity. *Journal of Experimental Psychology*, *84*(3), 483-486. doi:10.1037/h0029327

Helson, H. (1964). *Adaptation level theory*. New York, NY: Harper & Row.

Henry, L. A., & MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *American Journal on Mental Retardation*, *107*(6), 421-432.

Hinson, J. M., & Malone, C. J. Jr. (1980). Local contrast and maintained generalization. *Journal of the Experimental Analysis of Behavior*, *34*(3), 263-272. doi:10.1901/jeab.1980.34-263

Honig, W. K., & Urcuioli, P. J. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior, 36*, 404-445. doi:10.1901/jeab.1981.36-405

Horner, R. H., Albin, R. W., & Ralph, G. (1986). Generalization with precision: The role of negative teaching examples in the instruction of generalized grocery item selection. *Journal of the Association for Persons with Severe Handicaps, 11*(4), 300-308. doi:10.1177/154079698601100411

Horner, R. H., Bellamy, G. T., & Colvin, G. T. (1984). Responding in the presence of nontrained stimuli: Implications of generalization error patterns. *Journal of the Association for Persons with Severe Handicaps, 9*(4), 287-295. doi:10.1177/154079698400900406

Horner, R. H., Jones, D. N., & Williams, J. A. (1985). A functional approach to teaching generalized street crossing. *Journal of the Association for Persons with Severe Handicaps*, *10*, 71-78. doi:10.1177/154079698501000201

Horner, R. H., Sprague, J., & Wilcox, B. (1982). General case programming for community activities. In B. Wilcox & G. T. Bellamy (Eds.). *Design of high school programs for severely handicapped students* (pp. 61-98). Baltimore: Paul H. Brookes.

Horner, R. H., Williams, J. A., & Steveley, J. D. (1987). Acquisition of generalized telephone use by students with moderate and severe mental retardation. *Research in Developmental Disabilities, 8*, 229-247. doi:10.1016/0891-4222(87)90006-0

House, B., & Zeaman, D. (1962). Reversal and non-reversal shifts in discrimination learning in retardates. *Journal of Experimental Psychology*, *63*(5), 444. doi:10.1037/h0041656

Howard, R. W. (2000). Generalization and transfer: An interrelation of paradigms and a taxonomy of knowledge extension processes. *Review of General Psychology, 4*(3), 211-237. doi:10.1037//1089-2680.4.3.211

Howard, W. W. (1979). Maintained generalization gradients with contingent and non-contingent reinforcement. *Psychological Reports*, *45*, 275-279. doi:10.2466/pr0.1979.45.1.275

Hughes, C., Kaplan, L., Bernstein, R., Boykin, M., Reilly, C., Brigham, N., … Harvey, M. (2012). Increasing social interaction skills of secondary school students with autism and/or intellectual disability: A review of interventions. *Research & Practice for Persons with Severe Disabilities, 37*(4), 288-307. doi:10.2511/027494813805327214

Jarrold, C., & Baddeley, A. D. (1997). Short-term memory for verbal and visuo-spatial information in Down's syndrome. *Cognitive Neuropsychiatry, 45*, 531-544. doi:10.1080/135468097396351

Jarrold, C., Baddeley, A. D., & Phillips, C. E. (2002). Verbal short-term memory in Down syndrome: A problem of memory, audition, or speech? *Journal of Speech, Language, and Hearing Research, 45*, 531-544. doi:10.1044/1092-4388(2002/042)

Jarrold, C., Phillips, C., & Baddeley, A. D. (2007). Binding of visual and spatial short-term memory in Williams syndrome and moderate learning disability. *Developmental Medicine & Child Neurology*, *49*(4), 270-273. doi:10.1111/j.1469-8749.2007.00270.x

Jenkins, H. M., & Harrison, R. H. (1960). Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology, 59*, 246-253. doi:10.1037/h0041661

Johnston, S., Reichle, J., Feeley, K., & Jones, E. (2012). *Augmentative and alternative communication strategies for individuals with moderate to severe disabilities*. Baltimore, MD: Paul H. Brookes.

Kaiser, A. P., & Roberts, M. Y. (2013). Parent-implemented enhanced milieu teaching with preschool children who have intellectual disabilities. *Journal of Speech, Language, and Hearing Research*, 56(1), 295-309. doi:10.1044/1092-4388(2012/11-0231)

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings.* (2nd ed.). New York, NY: Oxford University Press.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology*. New York, NY: Appleton-Century-Crofts.

Kimble, G. A. (1961). *Hilgard and Marquis' conditioning and learning.* (2nd ed.). New York, NY: Appleton-Century-Crofts.

Koyama, T., & Wang, H-T. (2011). Use of activity schedule to promote independent performance of individuals with autism and other intellectual disabilities: A review. *Research in Developmental Disabilities, 32*, 2235-2242. doi:10.1016/j.ridd.2011.05.003

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single case design technical documentation*. Retrieved from What Works Clearinghouse website: https://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_scd.pdf

Kreibich, S. R., Chen, M., & Reichle, J. (2015). Teaching a child with autism to request breaks while concurrently increasing task engagement. *Language, Speech, and Hearing Services in Schools*, *46*, 256-265. doi:10.1044/2015_LSHSS-14-0081

Kuhn, D. E., Chirighin, A. E., & Zelenka, K. (2010). Discriminated functional communication: A procedural extension of functional communication training. Journal of Applied Behavior Analysis, 43(2), 249-264. doi:10.1901/jaba.2010.43-249

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

LaBerge, D. (1961). Generalization gradients in a discrimination situation. *Journal of Experimental Psychology, 62*(1), 88-94. doi:10.1037/h0043035

Lane, H., & Curran, C. (1963). Gradients of auditory generalization for blind, retarded children. *Journal of the Experimental Analysis of Behavior*, *6*(4), 585-588.

Laraway, S., Snycerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: Some further refinements. *Journal of Applied Behavior Analysis, 36*(3), 407-414. doi:10.1901/jaba.2003.36-407

Livesey, E. J., & McLaren, I. P. L. (2009). Discrimination and generalization along a simple dimension: Peak shift and rule-governed responding. *Journal of Experimental Psychology, 35*(4), 554-565. doi:10.1037/a0015524

Lovaas, O, I., Schreibman, L., Koegel, R., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology, 77*(3), 211-222. doi:10.1037/h0031015

Lyons, J., Klipec, W. D., & Steinsultz, G. (1973). The effect of chloropromazine on discrimination performance and the peak shift. *Physiological Psychology, 1*, 121-124. doi:10.3758/bf03326882

Malone, J. C., & Staddon, J. E. R. (1973). Contrast effects in maintained generalization gradients. *Journal of the Experimental Analysis of Behavior*, *19*(1), 167-179. doi:10.1901/jeab.1973.19-167

Marcell, M. M., & Armstrong, V. (1982). Auditory and visual sequential memory of Down syndrome and nonretarded children. *American Journal of Mental Deficiency*, *210*, 207-210.

Marcell, M. M., & Weeks, S. L. (1988). Short-term memory difficulties and Down's syndrome. *Journal of Mental Deficiency Research, 32*, 153-162. doi:10.1111/j.1365-2788.1988.tb01400.x

Marsh, G. (1972). Prediction of the peak shift in pigeons from gradients of excitation and inhibition. *Journal of Comparative and Physiological Psychology, 81*, 262-266. doi:10.1037/h0033538

McIlvane, W. J., Dube, W. V., Kledaras, J. B., Iennaco, F. M., & Stoddard, L. T. (1990). Teaching relational discrimination to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation, 95*, 283-296.

Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior, 37*, 149-155. doi:10.1901/jeab.1982.37-149

Michael, J. (1993). *Establishing operations*. *The Behavior Analyst, 16*, 191-206.

Michael, J. (2000). Implications and refinements of the establishing operation concept. *Journal of Applied Behavior Analysis, 33*, 401-410. doi:0.1901/jaba.2000.33-401

Morse, T. E., & Schuster, J. W. (1996). Grocery shopping skills for persons with moderate to profound intellectual disabilities: A review of the literature. *Education and Treatment of Children*, *19*, 487-517.

Mullen, E. M. (1995). *Mullen scales of early learning: AGS edition (MSEL-AGS)*. Circle Pines, MN: American Guidance Service.

Murphy, C., & Barnes-Holmes, D. (2010). Establishing five derived mands in three adolescent boys with autism. *Journal of Applied Behavior Analysis*, *43*(3), 537-541.

Nelson, K. R., & Wasserman, E. A. (1978). Temporal factors influencing the pigeon's successive matching-to-sample performance: Sample duration, intertribal interval, and retention interval. *Journal of Experimental Analysis of Behavior, 30*(2), 153-162. doi:10.1901/jeab.1978.30-153

Neurobehavioral Systems. (2004). *Presentation® software*. Berkeley, CA: Neurobehavioral Systems, Inc. Retrieved from https://www.neurobs.com/

Newhall, S. M., Burnham, R. W., & Clark, J. R. (1957). Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America, 47*(1), 43-56. doi:10.1364/josa.47.000043

Nicholson, J. N. & Gray, J. A. (1971). Behavioral contrast and peak shift in children. *British Journal of Psychology, 62*, 367-373. doi:10.1111/j.2044-8295.1971.tb02047.x

O'Donnell, J., & Saunders, K. J. (2003). Equivalence relations in individuals with language limitations and mental retardation. *Journal of the Experimental Analysis of Behavior, 80*(1), 131-157. doi:10.1901/jeab.2003.80-131

O'Neill, R. E., Faulkner, C., & Horner, R. H. (2000). The effects of general case training of manding responses on children with severe disabilities. *Journal of Developmental and Physical Disabilities*, 12(1), 43-60. doi:10.1023/A:1009456210739

O'Neill, R. E., & Reichle, J. (1993). Addressing socially motivated challenging behavior by establishing communicative alternatives: Basics of a general-case approach. In J. Reichle & D. Wacker (Eds.). *Communicative alternatives to challenging behavior: Integrating functional assessment and intervention strategies* (pp. 205-236). Baltimore, MD: Paul H. Brookes.

Orr, A. C., & Mast, M. (2014). Tablet-based communication and children with multiple disabilities: Lessons from the clinical setting. *Procedia – Social and Behavioral Sciences, 141*, 138-142. doi:10.1016/j.sbspro.2014.05.025

Osnes, P. G., & Lieblein, T. (2003). An explicit technology of generalization. *The Behavior Analyst Today, 3*(4), 364-374. doi:10.1037/h0099994

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Clarendon Press.

Pickering, S. J., & Gathercole, S. E. (2004). Distinctive working memory profiles in children with special educational needs. *Educational Psychology*, *24*(3), 393-408. doi:10.1080/0144341042000211715

Pierce, W. D., & Cheney, C. D. (2008). *Behavior analysis and learning*. New York, NY: Psychological Press.

Pierrel, R., & Sherman, J. G. (1960). Generalization of auditory intensity following discrimination training. *Journal of the Experimental Analysis of Behavior*, *3*, 313-322. doi:10.1901/jeab.1960.3-313

Powell, R. A., Symbaluk, D. G., & Honey, P. L. (2009). *Introduction to learning and behavior*. (3rd ed.). Belmont, CA: Wadsworth.

Powell, R. A., Symbaluk, D. G., & MacDonald, S. E. (2002). *Introduction to learning and behavior*. Belmont, CA: Wadsworth.

Purtle, R. B. (1973). Peak shift: A review. *Psychological Bulletin, 80*, 408-421. doi:10.1037/h0035233

Reichle, J., & Johnston, S. S. (1999). Teaching the conditional use of communicative requests to two school-age children with severe developmental disabilities. *Language, Speech, and Hearing Services in Schools, 30*, 324-334. doi:10.1044/0161-1461.3004.324

Reichle, J., & Wilkinson, K. M. (2012). Teaching learners to correspond graphic symbols to objects and events. In S. S. Johnston., J. Reichle., K. M. Feeley & E. A. Jones (Eds.), AAC strategies for individuals with moderate to severe disabilities (pp. 207-208). Baltimore, MD: Paul H. Brookes.

Rilling, M. (1977). Stimulus control and inhibitory processes. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 432-480). Englewood Cliffs, NJ: Prentice Hall.

Risley, T. (1964). Generalization gradients following two-response discrimination training. *Journal of the Experimental Analysis of Behavior*, *7*(2), 199-204.

Rodgers, J. P., & Thomas, D. R. (1982). Task specificity in nonspecific transfer and in extradimensional stimulus generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Process, 8*(4), 301-312. doi:10.1037/0097-7403.8.4.301

Rodriguez-Carmona, M., O'Neill-Biba, M., & Barbur, J. L. (2012). Assessing the severity of color vision loss with implications for aviation and other occupational environments. *Aviation, Space, and Environmental Medicine*, *83*(1), 19-29. doi:10.3357/ASEM.3111.2012

Sargisson, R. J., & White, K. G. (2001). Generalization of delayed matching to sample following training at different delays. *Journal of the Experimental Analysis of Behavior*, *75*(1), 1-14. doi: 10.1901/jeab.2001.75-1

Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior, 52*, 1-12. doi:10.1901/jeab.1989.52-1

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior, 54*, 239-250. doi:10.1901/jeab.1990.54-239

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior, 60*, 571-585. doi:10.1901/jeab.1993.60-571

Scavio, M. J. J., & Gormezano, I. (1974). CS intensity effects on rabbit nictitating membrane, conditioning, extinction and generalization. *Pavlovian Journal of Biological Science*, *9*, 25-34.

Schuchardt, K., Gebhardt, M., & Mäehler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research*, *54*(4), 346-353.

Schuster, R. H., & Gross, C. G. (1969). Maintained generalization gradients in the monkey. *Psychonomic Science*, *14*(5), 215-217. doi: 10.3758/bf03332799

Scruggs, T. E., & Mastropieri, M. A. (1984). Issues in generalization: Implications for special education. *Psychology in the Schools, 21*, 398-403. doi:10.1002/1520-6807(198407)21:3<397::aid-pits2310210321>3.0.co;2-7

Serna, R. W., Dube, W. V., & McIlvane, W. J. (1997). Assessing same/different judgments in individuals with severe intellectual disabilities: A status report. *Research in Developmental Disabilities, 18*(5), 343-368. doi:10.1016/s0891-4222(97)00015-2

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323. doi:10.1126/science.3629243

Sidman, M. (1969). Generalization gradients and stimulus control in delayed matching-to-sample. *Journal of the Experimental Analysis of Behavior, 12*, 745-757. doi:10.1901/jeab.1969.12-745

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research*, *14*, 5-13. doi:10.1044/jshr.1401.05

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213-245). Hillsdale, NJ: Erlbaum.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127-146. doi:10.1901/jeab.2000.74-127

Sigafoos, J., & Roberts-Pennell, D. (1999). Wrong-item format: A promising intervention for teaching socially appropriate forms of rejecting to children with developmental disabilities? *Augmentative and Alternative Communication, 15*, 135-140. doi:10.1080/07434619912331278635

Snell, M. E., Brady, N., McLean, L., Ogletree, B. T., Siegel, E., Sylvester, L., … Sevcik, R. (2010). Twenty years of communication intervention research with individuals who have severe intellectual and developmental disabilities. *American Journal of Intellectual and Developmental Disabilities, 115*(5), 364-380. doi:10.1352/1944-7558-115-5.364

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales, second edition (Vineland-II): The parent/caregiver rating form*. Livonia, MN: Pearson Assessment.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44*, 430-444. doi:10.1037/h0062885

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, *77*, 556-566.

Sprague, J. R., & Horner, R. H. (1984). The effects of single instance, multiple instance and general case training on generalized vending machine use by moderately and severely handicapped students. *Journal of Applied Behavior Analysis, 17*, 273-278. doi:10.1901/jaba.1984.17-273

Steere, D. E., Strauch, J. D., Powell, T. H., & Butterworth, J. (1990). Promoting generalization from a teaching setting to a community-based setting among persons with severe disabilities: A general case programming approach. *Education and Treatment of Children*, *13*(1), 5-20.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349-367. doi:10.1901/jaba.1977.10-349

Stokes, J. V., Cameron, M. J., Dorsey, M. F., & Fleming, E. (2004). Task analysis, correspondence training, and general case instruction for teaching personal hygiene skills. *Behavioral Interventions, 19*(2), 121-135. doi:10.1002/bin.153

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20*, 337-355. doi:10.1016/s0005-7894(89)80054-1

Stromer, R., & Stromer, J. B. (1990). The formation of arbitrary stimulus classes in matching to complex samples. *Psychological Record*, *40*, 51-66.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York, NY: Academic Press, Inc.

Taft, R. J., & Mason, L. H. (2011). Examining effects of writing interventions: Highlighting results for students with primary disabilities other than learning disabilities. *Remedial and Special Education, 32*(5), 359-370. doi:10.1177/0741932510362242

Tennison, L. R., & Hinson, J. M. (1993). Range effects in transient and maintained generalization gradients. *Behavioral Processes*, *29*, 201-216. doi:10.1016/0376-6357(93)90124-a

Terrace, H. S. (1966). Stimulus control. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 271-344). New York: Appleton-Century-Crofts.

Thomas, D. R. (1993). A model of adaptation-level effects on stimulus generalization. *Psychological Review, 100*, 658-673. doi:10.1037/0033-295x.100.4.658

Thomas, D. R., & Curran, P. J. (1988). The role of response directedness in discriminative and conditional stimulus control. *Bulletin of the Psychonomic Society*, *26*(4), 378-380. doi:10.3758/BF03337688

Thomas, D. R., Lusky, M., & Morrison, S. (1992). A comparison of generalization functions and frame of reference effects in different training paradigms. *Perception & Psychophysics*, *51*(6), 529-540. doi:10.3758/bf03211650

Thomas, D. R., McKelvie, A. R., Ranney, M., & Moye, T. B. (1981). Interference in pigeons' long-term memory viewed as a retrieval problem. *Animal Learning & Behavior*, *9*(4), 581-586. doi:10.3758/bf03209794

Thomas, D. R., Mood, K., Morrison, S., & Wiertelak, E. (1991). Peak shift revisited: A test of alternative interpretations. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 130-140. doi:10.1037//0097-7403.17.2.130

Tomporowski, P. D., & Tinsley, V. (1997). Attention in mentally retarded persons. In W. E. MacLean, Jr. (Ed.), *Ellis' handbook of mental deficiency, psychological theory and research* (3rd ed.) (pp. 219-244). Mahwah, NJ: Erlbaum.

Tiger, J. H., Hanley, G. P., & Heal, N. A. (2006). The effectiveness of and preschoolers' preferences for variations of multiple schedule arrangements. *Journal of Applied Behavior Analysis, 39*(4), 475-488. doi:10.1901/jaba.2006.48-06

Morse, T. E., & Schuster, J. W. (1996). Grocery shopping skills for persons with moderate to profound intellectual disabilities: A review. *Education & Treatment of Children, 19*(4), 487-516.

Urcuioli, P. J. (2011). Emergent identity matching after successive matching training I: Reflexivity or generalized identity. *Journal of Experimental Analysis of Behavior, 96*(3), 329-341. doi:10.1901/jeab.2011.96-329

Urcuioli, P. J., & Swisher, M. (2012). Emergent identity matching after successive matching training II: Reflexivity or transitivity. *Journal of Experimental Analysis of Behavior, 97*(1), 5-27. doi:10.1901/jeab.2012.97-5

Vervliet, B., Iberico, C., Vervoort, E., & Baeyens, F. (2011). Generalization gradients in human predictive learning: Effects of discrimination and within-subjects testing. *Learning and Motivation, 42*, 210-220. doi:10.1016/j.lmot.2011.03.004

Waggoner, T. L. (1994). *The color vision testing made easy® test* (CVTMET). Elgin, IL: Good-Lite Co.

Waggoner, T. L. (2011). Ishihara color vision test, 24 plate edition. Elgin, IL: Good-Lite Co.

Wasserman, E. A. (1976). Successive matching-to-sample in the pigeon: Variations on a theme by Konorski. *Behavioral Research Methods & Instrumentation, 8*(3), 278-282. doi:10.3758/bf03201713

Watson, L. S. (1966). *The relationship between discrimination and stimulus generalization in severely retarded children*. Unpublished doctoral dissertation. The Ohio State University, Columbus, Ohio.

Wheatley, K. L., & Thomas, D. R. (1974). Relative and absolute density of reinforcement as factors influencing the peak shift. *Journal of the Experimental Analysis of Behavior, 22*, 409-418. doi:10.1901/jeab.1974.22-409

White, K. G., & Thomas, D. R. (1979). Topographically tagged stimulus control: Maintained generalization and stimulus-specific gradients. *Bulletin of the Psychonomic Society*, *13*(5), 275-278. doi:10.3758/bf03336869

Wilhelm, H., & Lovaas, O. I. (1976). Stimulus overselectivity: A common feature in autism and mental retardation. *American Journal on Mental Deficiency. 81*(1), 26-31. doi:10.1352/0895-8017(2006)111[447:maosoi]2.0.co;2

Wilkie, D. M. (1972). The peak shift and behavioral contrast: Effects of discrimination training with delayed reinforcement. *Psychonomic Science, 26*, 257-258. doi:10.3758/bf03328610

Wills, S., & Mackintosh, N. J. (1998). Peak shift on an artificial dimensions. *Quarterly Journal of Experimental Psychology, 51B*, 1-31.

Winograd, E. (1965). Maintained generalization testing of conditioned suppression. *Journal of the Experimental Analysis of Behavior*, *8*(1), 47-51. doi:10.1901/jeab.1965.8-47

Wisniewski, M. G., Church, B. A., & Mercado III, E. (2009). Learning-related shifts in generalization gradients for complex sounds. *Learn Behavior, 37*(4), 325-335. doi:10.3758/lb.37.4.325

Woolcock, W. W., Lyon, S. R., & Woolcock, K. P. (1987). General case simulation instruction and the establishment and maintenance of work performance. *Research in Developmental Disabilities, 8*, 427-447. doi:10.1016/0891-4222(87)90024-2

Yarczower, M., Dickson, J. F., & Gollub, L. R. (1966). Some effects on generalization gradients of tandem schedules. *Journal of the Experimental Analysis of Behavior*, 9(6), 631-639. doi:10.1901/jeab.1966.9-631

Zaine, I., Domeniconi, C., & de Rose, J. C. (2014). Simple and conditional discrimination and specific reinforcement in teaching reading: An intervention package. *The Analysis of Verbal Behavior*, *30*(2), 193-204. doi:10.1007/s40616-014-0010-2

Zeiler, M. D. (1969). Repeated measurements of reinforcement schedule effects on gradients of stimulus control. *Journal of the Experimental Analysis of Behavior*, 12(3), 451-461. doi:10.1901/jeab.1969.12-451

Zoeke, B., Sarris, V., & Hofer, G. (1988). Psychophysical context effects in chickens (*Gallus gallus domesticus*). *International Journal of Comparative Psychology*, *1*, 167-178.

Zygmont, D. M., Lazar, R. M., Dube, W. V., & McIlvane, W. J. (1992). Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: A methodological note. *Journal of the Experimental Analysis of Behavior*, *57*, 109-117. doi:10.1901/jeab.1992.57-109

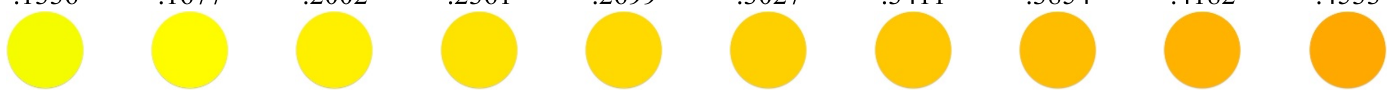
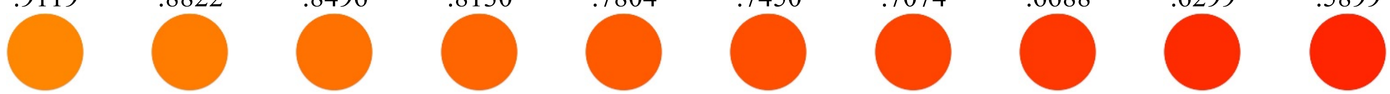
Table 1

Definitions of Generalization Strategies (Stokes & Baer, 1977)

Strategy	Definition
Train and hope	Behavior change is effected under limited conditions; assessing for generalization but not programming for it.
Sequential modification	Behavior change is effected under limited conditions; assessing for generalization and directly training in "generalization" settings if no or limited generalization occurs.
Introduce to natural maintaining contingencies	Teach the skill to be reinforced by naturally existing contingencies.
Train sufficient exemplars	Teaching the student to respond to a subset of all possible stimulus and response examples and then assess generalization on untrained examples.
Train loosely	Teaching is conducted with relatively little control over the stimuli presented and the correct response allowed.
Use indiscriminable contingencies	Making the contingencies of reinforcement or punishment, or the setting events that signal the presence or absence of those contingencies indiscriminable.
Program common stimuli	Incorporating into intervention the stimulus components that occur in common in both the intervention and generalization settings.
Mediate generalization	Incorporating tools or strategies that the learner can readily use across situations (e.g., self-management procedures).
Train to generalize	Reinforcing generalization itself as if it were an explicit behavior.

Table 2

Three Sets of Colors Implemented in Pilot Study

Color Set	Color Stimuli									
Blue to green										
Wavelength (nm)	470	477	483	489	493	496	499	502	504	506
Luminosity Value[2]	.1336	.1677	.2002	.2361	.2699	.3027	.3411	.3854	.4182	.4533
Yellow to orange										
Wavelength (nm)	577	580	584	588	591	594	597	600	603	606
Luminosity Value	.9119	.8822	.8496	.8130	.7804	.7450	.7074	.6688	.6299	.5899
Orange to red										
Wavelength (nm)	616	619	622	625	628	631	634	637	641	645
Luminosity Value	.4542	.4151	.3784	.3426	.3061	.2709	.2386	.2094	.1754	.1458

[2] Luminosity reflects the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_{\lambda}}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_{λ} is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005). The original sensitivity values for each wavelength was found on http://web.archive.org/web/20070314050445/http://www.cvrl.org/database/data/lum/ssvl2e_1.txt

Table 3

Corresponding Luminosity Values and Intensity Values for Each Monochromatic Color Stimulus Used in the Pilot Study

Color Set	Wavelength (in nm)	Luminosity Value[3]	RGB (Intensity Values)[4]
Blue-green stimuli set (range: 36 nm)	470	.1336	(0, 153, 255)
	477	.1677	(0, 188, 255)
	483	.2002	(0, 219, 255)
	489	.2361	(0, 249, 255)
	493	.2699	(0, 255, 216)
	496	.3027	(0, 255, 178)
	499	.3411	(0, 255, 140)
	502	.3854	(0, 255, 102)
	504	.4182	(0, 255, 76)
	506	.4533	(0, 255, 51)
Yellow-orange stimuli set (range: 29 nm)	577	.9119	(244, 255, 0)
	580	.8822	(255, 255, 0)
	584	.8496	(255, 239, 0)
	588	.8130	(255, 223, 0)
	591	.7804	(255, 211, 0)
	594	.7450	(255, 200, 0)
	597	.7074	(255, 188, 0)
	600	.6688	(255, 176, 0)
	603	.6299	(255, 164, 0)
	606	.5899	(255, 153, 0)
Orange-red stimuli set (range: 28 nm)	616	.4542	(255, 113, 0)
	619	.4151	(255, 102, 0)
	622	.3784	(255, 90, 0)
	625	.3426	(255, 78, 0)
	628	.3061	(255, 66, 0)
	631	.2709	(255, 54, 0)
	634	.2386	(255, 43, 0)
	637	.2094	(255, 31, 0)
	641	.1754	(255, 15, 0)
	645	.1458	(255, 0, 0)

[3] Luminosity reflects the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_\lambda}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_λ is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = \ 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005).

[4] RGB values refer to the Red-Green-Blue intensity values and for each primary color the value ranges from 0 to 255. With this system, 16,777,216 (256^3) discrete combinations of R, G and B values are allowed, providing millions of different (though not necessarily distinguishable) colors.

Table 4

Percent of Independent Selections for Each Target Response in Each of Two Forced Choice Response Modes

Stimulus Values (in nm)	Forced Choice Between Colors (%)		Forced Choice Between Yes and No (%)				95% CI	
	M	*SD*	*M*	*SD*	*t*(29)	*p*	*LL*	*UL*
Blue to Green Color Set (Target Response: "Green")								
470	2	8	0	0	-1.44	.16	-1	5
477	0	0	2	8	-1.44	.16	-5	1
483	0	0	0	0	N/A	N/A	N/A	N/A
489	1	6	0	0	1.00	.33	-1	3
493	0	0	1	6	-1.00	.33	-3	1
496	17	30	11	24	1.15	.26	-4	15
499	77	31	66	40	1.72	.10	-2	25
502	99	6	99	6	0	1.00	-3	3
504	100	0	99	6	1.00	.33	-1	3
506	99	6	100	0	-1.00	.33	-3	1
Yellow to Orange Color Set (Target Response: "Orange")								
577	7	25	16	36	-1.76	.09	-19	1
580	8	26	18	34	-1.97	.06	-20	0
584	8	23	16	35	-1.76	.09	-17	1
588	9	19	19	36	-2.07	.048	-20	-0.1
591	22	34	30	37	-1.23	.23	-21	5
594	38	41	49	43	-1.98	.06	-23	0
597	59	42	67	39	-1.05	.30	-23	7
600	79	27	86	29	-1.22	.23	-18	4
603	88	27	90	20	-0.57	.57	-10	6
606	91	23	91	19	0	1.00	-9	9
Orange to Red Color Set (Target Response: "Red")								
616	0	0	0	0	N/A	N/A	N/A	N/A
619	0	0	0	0	N/A	N/A	N/A	N/A
622	2	8	8	23	-1.54	.13	-13	2
625	6	15	16	29	-2.34	.03	-19	-1
628	17	31	29	38	-1.88	.07	-25	1
631	51	43	49	43	0.22	.83	-14	17
634	71	35	69	41	0.35	.73	-11	16
637	97	13	90	25	1.53	.14	-2	16
641	100	0	100	0	N/A	N/A	N/A	N/A
645	100	0	100	0	N/A	N/A	N/A	N/A

Note. M = mean; *SD* = standard deviation; CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table 5

Response Latency in Each of Two Forced Choice Response Modes

Stimulus Values (in nm)	Forced Choice Between Colors (s)		Forced Choice Between Yes and No (s)				95% CI	
	M	*SD*	*M*	*SD*	*t*(29)	*p*	*LL*	*UL*
Blue to Green Color Set								
470	0.83	0.31	0.89	0.27	-0.89	.38	-0.19	0.07
477	0.89	0.37	0.79	0.22	1.51	.14	-0.03	0.22
483	0.83	0.33	0.84	0.24	-0.20	.84	-0.13	0.10
489	0.92	0.37	0.86	0.29	0.92	.37	-0.08	0.20
493	0.91	0.31	0.93	0.28	-0.29	.77	-0.14	0.11
496	1.34	0.54	1.18	0.50	2.33	.03	0.02	0.30
499	1.55	0.95	1.33	0.57	1.34	.19	-0.12	0.55
502	0.97	0.32	1.04	0.32	-1.16	.26	-0.20	0.06
504	0.86	0.31	0.94	0.33	-1.18	.25	-0.23	0.06
506	0.81	0.23	0.84	0.25	-0.64	.53	-0.12	0.06
Yellow to Orange Color Set								
577	0.81	0.27	0.90	0.32	-1.41	.17	-0.22	0.04
580	0.81	0.26	0.87	0.41	-0.77	.45	-0.25	0.11
584	0.94	0.43	0.92	0.45	0.19	.85	-0.18	0.21
588	1.05	0.45	1.03	0.48	0.16	.87	-0.16	0.19
591	1.24	0.55	1.25	0.49	-0.17	.87	-0.23	0.20
594	1.12	0.36	1.37	0.55	-2.50	.02	-0.46	-0.05
597	1.22	0.44	1.33	0.58	-1.10	.28	-0.31	0.09
600	1.14	0.40	1.09	0.42	0.53	.60	-0.14	0.23
603	1.00	0.42	1.11	0.44	-1.24	.22	-0.28	0.07
606	0.89	0.32	1.00	0.41	-1.51	.14	-0.26	0.04
Orange to Red Color Set								
616	0.77	0.25	0.90	0.39	-1.71	.10	-0.27	0.02
619	0.81	0.31	0.95	0.49	-1.69	.10	-0.31	0.03
622	0.90	0.42	0.99	0.51	-1.48	.15	-0.23	0.04
625	0.96	0.42	1.01	0.52	-0.57	.57	-0.22	0.12
628	1.12	0.50	1.15	0.51	-0.22	.83	-0.26	0.21
631	1.30	0.55	1.25	0.46	0.51	.61	-0.16	0.26
634	1.21	0.47	1.12	0.43	0.89	.38	-0.11	0.28
637	0.95	0.53	1.12	0.68	-1.21	.24	-0.48	0.12
641	0.74	0.23	0.79	0.22	-0.95	.35	-0.16	0.06
645	0.73	0.27	0.81	0.24	-1.57	.13	-0.19	0.02

Note. *M* = mean; *SD* = standard deviation; CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Table 6

Outcomes of Within-Subject Two-Way Analysis of Variances (ANOVAs) for Percent of Independent Selections of the Target Choice Given the PSI Stimulus for Each Color Set in the Pilot Study

Effect	*SS*	*df*	*F*	*p*
Between subjects	10.11	29		
Response mode	0.004	1	0.045	0.833
Response mode × Subject	2.50	29		
Color set	2.587	2	6.168	0.004**
Color set × Subject	12.164	58		
Response mode × Color set	0.392	2	3.446	0.039*
Response mode × Color set × Subject	3.299	58		

Note. SS = sums of squares, * $p < .05$, ** $p < .01$.

Table 7

Outcomes of Within-Subject Two-Way Analysis of Variances (ANOVAs) Results for Response Latency Given the PSI Stimulus for Each Color Set in the Pilot Study

Effect	*SS*	*df*	*F*	*p*
Between subjects	22.21	29		
Response mode	0.002	1	0.006	0.936
Response mode × Subject	7.702	29		
Color set	1.252	2	1.793	0.176
Color set × Subject	20.249	58		
Response mode × Color set	1.719	2	3.856	0.027*
Response mode × Color set × Subject	12.924	58		

Note. SS = sums of squares. * $p < .05$.

Table 8

The Three Finalized Color Sets and the Residual Sum of Squares (RSS) Values Based on Comparisons of the Actual Data and Hypothetical Ideal Data Averaged Across Thirty Typically Developing Adults' Performance for Each of Two Forced Choice Response Modes for Each of Three Finalized Color Sets

Color Set	Stimulus Values (in nm)	Luminosity Values[5]	Forced Choice Between Colors		RSS Value	Forced Choice Between Yes and No		RSS Value	Average RSS Value
			Actual Data	Ideal Data		Actual Data	Ideal Data		
Blue to Green	483	0.2002	0	0	0.08	0	0	0.13	0.11
	489	0.2361	0.01	0		0	0		
	493	0.2699	0	0		0.01	0		
	496	0.3027	0.17	0		0.11	0		
	499	0.3411	0.77	1		0.66	1		
	502	0.3854	0.99	1		0.99	1		
	504	0.4182	1.00	1		0.99	1		
	506	0.4533	0.99	1		1	1		
	508	0.4904	1.00	1		1	1		
Yellow to Orange	580	0.8822	0.08	0	0.69	0.18	0	0.59	0.64
	584	0.8496	0.08	0		0.16	0		
	588	0.8130	0.09	0		0.19	0		
	591	0.7804	0.22	0		0.30	0		
	594	0.7450	0.38	1		0.49	1		
	597	0.7074	0.59	1		0.67	1		
	600	0.6688	0.79	1		0.86	1		
	603	0.6299	0.88	1		0.90	1		
	606	0.5899	0.91	1		0.91	1		
Orange to Red	619	0.4151	0	0	0.36	0	0	0.48	0.42
	622	0.3784	0.02	0		0.08	0		
	625	0.3426	0.06	0		0.16	0		
	628	0.3061	0.17	0		0.29	0		
	631	0.2709	0.51	1		0.49	1		
	634	0.2386	0.71	1		0.69	1		
	637	0.2094	0.97	1		0.90	1		
	641	0.1754	1	1		1	1		
	645	0.1458	1	1		1	1		

Note. On the finalized blue-to-green color set, the percent of independent selection of the color stimulus of 508 nm was assumed as 1.00, which was not directly examined in the pilot study but inferred from the responding to the color stimulus of 506 nm and the color stimulus of 504 nm.

[5] Luminosity reflects the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_\lambda}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_λ is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\,nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005). The original sensitivity values for each wavelength was found on http://web.archive.org/web/20070314050445/http://www.cvrl.org/database/data/lum/ssvl2e_1.txt

Table 9

An Example of the Teaching Stimulus Locations and Test Stimuli for Each of Three Intervention Conditions in the Primary Experiment

Experimental Condition	Color Set	Color Property	Stimuli Within Stimulus Class B				PSI	Stimuli Within Stimulus Class A			
			4	3	2	1		1	2	3	4
Simple discrimination	Blue-to-green	Wavelength (nm)	483	489	493	496	499	502	504	506	508
		Luminosity Value[6]	.2002	.2361	.2699	.3027	.3411	.3854	.4182	.4533	.4904
		Teaching stimulus					S_A				
Maximal-difference	Yellow-to-orange	Wavelength (nm)	580	584	588	591	594	597	600	603	606
		Luminosity Value	.8822	.8496	.8130	.7804	.7450	.7074	.6688	.6299	.5899
		Teaching stimuli	S_B				S_A				
Minimal-difference	Orange-to-red	Wavelength (nm)	619	622	625	628	631	634	637	641	645
		Luminosity Value	.4151	.3784	.3426	.3061	.2709	.2386	.2094	.1754	.1458
		Teaching stimuli			S_B		S_A				

Note. PSI refers to the stimulus at the point of subjective indifference, which refers to the value for which two responses are equally likely (Thomas et al., 1992). S_A and S_B refer to the teaching stimuli for the two conditional discrimination conditions, and there is only S_A in the simple discrimination condition. S_A and S_B were allocated as right and left responses (from the view of the participant's facing the laptop), respectively.

[6] Luminosity reflects the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_\lambda}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_λ is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005). The original sensitivity values for each wavelength was found on http://web.archive.org/web/20070314050445/http://www.cvrl.org/database/data/lum/ssvl2e_1.txt

Table 10

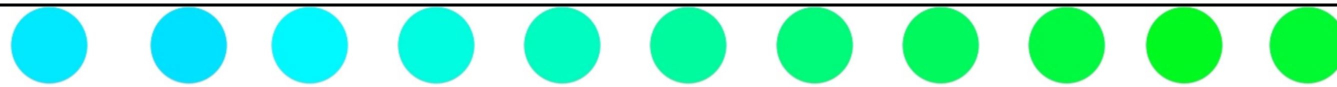
Participants Performance in Assessments for Color-Blindness, Receptive and Expressive Communication, Cognitive and Adaptive Behaviors

Participant	Color Blindness Test[a]	Receptive Communication Skills[b]	Expressive Communication Skills[c]	Cognitive Functioning Skills[d]	Adaptive Behavior Skills[e]
Bruce	Passed	Raw score = 13 Standard score <55 Percentile rank < 1st Age equivalent = 1:5	Raw score = 16 Standard score < 55 Percentile rank < 1st Age equivalent = 1:10	Subscale *T* score < 20 Age equivalent = 2:7	Standard score = 64 Percentile rank = 1st (at the low level)
Gabi	Passed	Raw score = 27 Standard score = 73 Percentile rank = 4th Age equivalent = 2:5	Raw score = 22 Standard score = 68 Percentile rank = 2nd Age equivalent = 2:3	Subscale *T* score < 24 Age equivalent = 2:6	Standard score = 94 Percentile rank = 34th (at the adequate level)
Bryan	Passed	Raw score = 26 Standard score = 76 Percentile rank = 5th Age equivalent = 2:7	Raw score = 6 Standard score < 55 Percentile rank < 1st Age equivalent = 1:3	Subscale *T* score ≤ 30 Age equivalent = 2:5	Standard score = 62 Percentile rank = 1st (at the low level)
James	Passed	Raw score = 14 Standard score = 55 Percentile rank < 1st Age equivalent = 1:6	Nonverbal; Not testable	Subscale *T* score < 20 Age equivalent = 1:9	Standard score = 65 Percentile rank = 1st (at the low level)
Corbin	Passed	Raw score = 37 Standard score = 79 Percentile = 8th Age equivalent = 3:3	Raw score = 30 Standard score = 73 Percentile = 4th Age equivalent = 2:11	Subscale *T* score ≤ 30 Age equivalent = 3:0	Standard score = 73 Percentile rank = 4th (at the moderately low level)

Note. a = Color blindness test was conducted using the Color Vision Testing Made Easy ® test (CVTMET); b = Receptive communication skills were assessed using the Receptive One-Word Picture Vocabulary Test, Third Edition (ROWPVT-3); c = Expressive communication skills were assessed using the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-3); d = Cognitive functioning skills were assessed using the Mullen Scales of Early Learning, AGS Edition (MSEL); e = Adaptive behavior skills were assessed using the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II).

Table 11

Color Samples from Each of Three Finalized Color Sets Including the "Perfect" Colors for Each Color Set in the Primary Experiment

Color Set	"Perfect" Colors	Color Stimuli as Samples									"Perfect" Colors
Blue to green											
						(PSI)					
Wavelength (nm)	485	483	489	493	496	499	502	504	506	508	507
Luminosity Value[7]	.2116	.2002	.2361	.2699	.3027	.3411	.3854	.4182	.4533	.4904	.4716
Yellow to orange											
						(PSI)					
Wavelength (nm)	582	580	584	588	591	594	597	600	603	606	604
Luminosity Value	.8653	.8822	.8496	.8130	.7804	.7450	.7074	.6688	.6299	.5899	.6167
Orange to red											
						(PSI)					
Wavelength (nm)	620	619	622	625	628	631	634	637	641	645	643
Luminosity Value	.4026	.4151	.3784	.3426	.3061	.2709	.2386	.2094	.1754	.1458	.1602

[7] Luminosity reflects the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_\lambda}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_λ is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005).

Table 12

Counterbalanced Order of Three Experimental Conditions Implemented for Each of the Five Participants with Moderate to Severe IDD

Color Set	Participants				
	Bruce	Gabi	Bryan	James	Corbin
Blue to Green	Minimal-difference conditional discrimination	Maximal-difference conditional discrimination	Minimal-difference conditional discrimination	Maximal-difference conditional discrimination	Simple discrimination
Yellow to Orange	Maximal-difference conditional discrimination	Minimal-difference conditional discrimination	Simple discrimination	Simple discrimination	Minimal-difference conditional discrimination
Orange to Red	Simple discrimination	Simple discrimination	Maximal-difference conditional discrimination	Minimal-difference conditional discrimination	Maximal-difference conditional discrimination

Table 13

Number of the Opportunities Divided by Total Opportunities in Which Invalid Responses[8] *Occurred During Baseline and Generalization Testing for Each of Five Participants with Moderate to Severe IDD*

Participant	Baseline			Generalization Testing		
	Single	Maximal	Minimal	Single	Maximal	Minimal
Bruce	0/45	2/45	1/45	2/45	0/45	1/45
			0/45			
Gabi	0/45	0/45	0/45	0/45	0/45	0/45
Bryan	0/45	0/45	0/45	0/45	0/45	1/45
James	2/45	0/45	1/45	0/45	0/45	0/45
Corbin	0/45	0/45	0/45	0/45	1/45	0/45

[8] Invalid responses refer to the participant pressing the two switches at the same time or the experimenter accidentally pressing the switch prior to the participant making a response. For those with two invalid responses, none of them were found to be for the same color sample. The number of 45 refers to the total responding opportunities during baseline, generalization testing, and maintenance for each experimental condition.

Table 14

The RSS Value for the Baseline Generalization Gradient in Each of Three Intervention Conditions for Each Participant

Participant	Condition	Color Set	RSS Value
Bruce	Simple	Orange-to-Red	2.32
	Maximal	Yellow-to-Orange	1.07
	Minimal	Blue-to-Green	0.89[a]
			1.04[b]
Gabi	Simple	Orange-to-Red	1.16
	Maximal	Blue-to-Green	0.80
	Minimal	Yellow-to-Orange	0.84
Bryan	Simple	Yellow-to-Orange	3.00
	Maximal	Orange-to-Red	4.04
	Minimal	Blue-to-Green	2.60
James	Simple	Yellow-to-Orange	2.89
	Maximal	Blue-to-Green	2.24
	Minimal	Orange-to-Red	3.24
Corbin	Simple	Blue-to-Green	0.44
	Maximal	Orange-to-Red	1.28
	Minimal	Yellow-to-Orange	0.80

Note. a = the initial baseline data for Bruce; b = the re-baseline data for Bruce.

Table 15

Response Reliability (Agreements/[Agreements + Disagreements] × 100%) for Each Experimental Condition for Each of Five Participants

Participant	Baseline			Intervention			Generalization Testing		
	Single	Maximal	Minimal	Single	Maximal	Minimal	Single	Maximal	Minimal
Bruce	5/5 100%	5/5 100%	5/5 100%	3/3 100%	6/6 100%	27/27 100%	5/5 100%	5/5 100%	10/10 100%
Gabi	5/5 100%	5/5 100%	5/5 100%	3/6 100%	3/3 100%	6/16 100%	5/5 100%	5/5 100%	0/5 N/A
Bryan	5/5 100%	5/5 100%	5/5 100%	13/16 100%	10/47 100%	12/31 100%	5/5 100%	0/5 N/A	5/5 100%
James	5/5 100%	5/5 100%	5/5 100%	2/9 100%	18/42 100%	10/20 100%	5/5 100%	5/5 100%	0/5 N/A
Corbin	5/5 100%	5/5 100%	5/5 100%	3/3 100%	10/17 100%	9/9 100%	5/5 100%	5/5 100%	5/5 100%

Table 16

Intervention Fidelity (Agreements/[Agreements + Disagreements] × 100%) for Each Experimental Condition for Each of Five Participants

Participant	Baseline			Intervention			Generalization Testing		
	Single	Maximal	Minimal	Single	Maximal	Minimal	Single	Maximal	Minimal
Bruce	5/5 100%	5/5 100%	5/5 100%	3/3 86.67%	6/6 86.67%	27/27 98.15%	5/5 97.78%	5/5 100%	10/10 100%
Gabi	5/5 100%	5/5 100%	5/5 100%	3/6 100%	3/3 100%	6/16 98.33%	5/5 93.33%	5/5 100%	0/5 N/A
Bryan	5/5 95.56%	5/5 100%	5/5 93.33%	13/16 97.69%	10/47 98%	12/31 96.67%	5/5 100%	0/5 N/A	5/5 100%
James	5/5 100%	5/5 100%	5/5 97.78%	2/9 100%	18/42 100%	10/20 100%	5/5 100%	5/5 100%	0/5 N/A
Corbin	5/5 100%	5/5 100%	5/5 97.78%	3/3 100%	10/17 99%	9/9 98.89%	4/5 100%	5/5 100%	0/5 N/A

Table 17

Percent of Correct Target Choice for Sample S_A, Mean Percent of Selection of the Target Choice for Samples From Each of Two Stimulus Classes (i.e., Stimulus Class B and Stimulus Class A) as Demonstrated in the Baseline and Generalization Testing Generalization Gradients for Each of Five Participants

Participant	Condition	Baseline			Generalization Testing		
		Sample S_A	Stimulus Class B	Stimulus Class A	Sample S_A	Stimulus Class B	Stimulus Class A
Bruce	Simple	40%	50%	55%	80%	31.25%	68.75%
	Maximal	60%	28.75%	75%	80%	40%	85%
	Minimal	80%[a]	20%[a]	62.5%[a]	50%	10%	95%
		60%[b]	15%[b]	65%[b]			
Gabi	Simple	40%	25%	85%	100%	35%	95%
	Maximal	80%	40%	95%	80%	20%	95%
	Minimal	100%	20%	75%	60%	15%	85%
Bryan	Simple	20%	45%	40%	20%	40%	80%
	Maximal	0%	55%	35%	60%	30%	90%
	Minimal	60%	50%	45%	80%	21.25%	100%
James	Simple	25%	42.5%	40%	80%	65%	85%
	Maximal	60%	40%	50%	20%	50%	50%
	Minimal	80%	73.75%	55%	60%	60%	40%
Corbin	Simple	40%	0%	90%	100%	0%	100%
	Maximal	20%	10%	70%	100%	45%	95%
	Minimal	80%	5%	60%	40%	20%	85%

Note. The target choice herein for samples from both stimulus Class A and stimulus Class B was the same as the one to which the sample S_A corresponded.

Table 18

Number of Teaching Opportunities to Mastery Criterion for Each of Three Acquisition Conditions for Each Participant and Averaged Across Five Participants

Participant	Simple Discrimination Condition	Maximal-Difference Conditional Discrimination	Minimal-Difference Conditional Discrimination
Bruce	30	60	150[a]; 120[b]
Gabi	60	30	160
Bryan	160	470	310
James	90	420	200
Corbin	30	170	90
Average	74	230	176

Note. a = the initial intervention data for Bruce; b = the modified intervention data for Bruce.

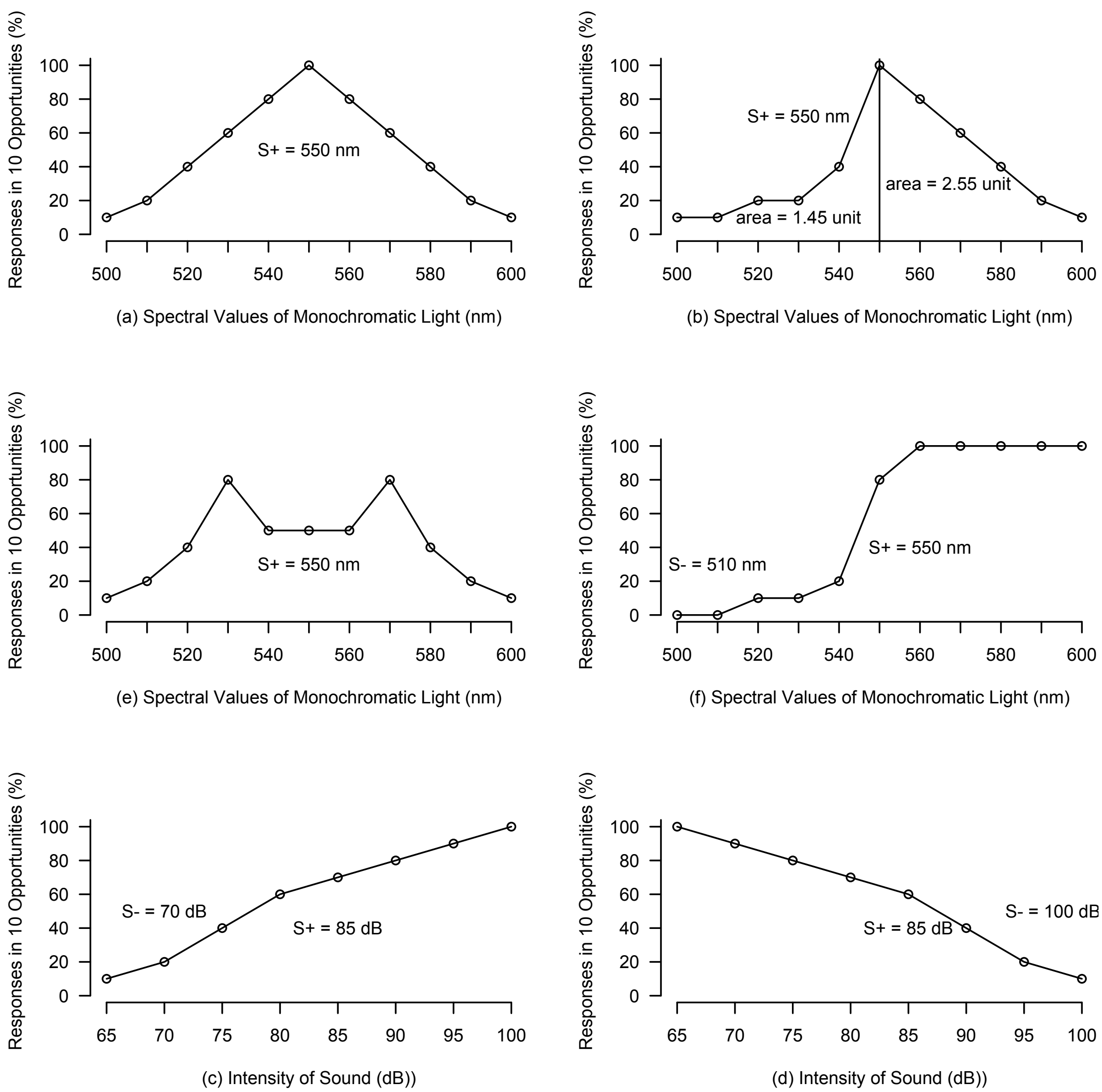


Figure 1. Examples of different generalization gradients.

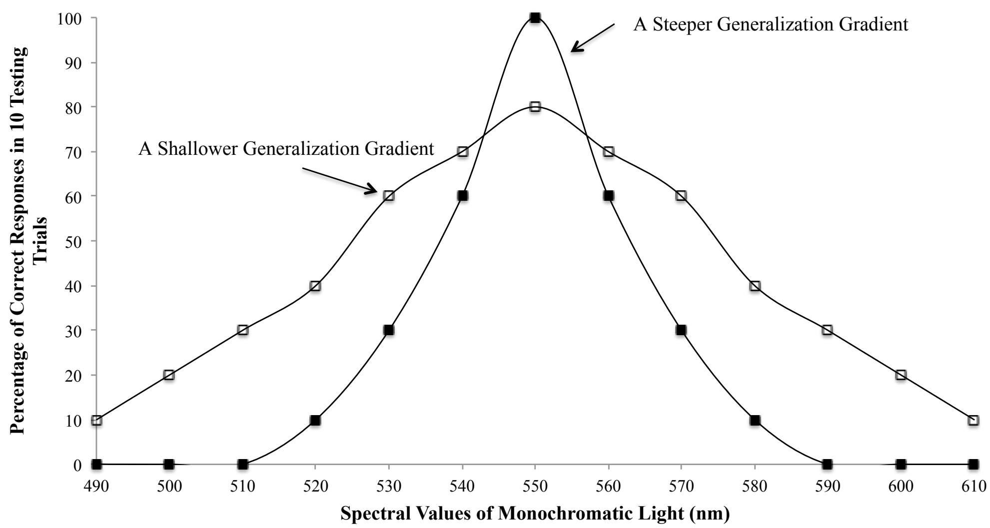


Figure 2. A hypothetical graph for a steeper generalization gradient and a shallower generalization gradient along the stimulus dimension of monochromatic light, respectively.

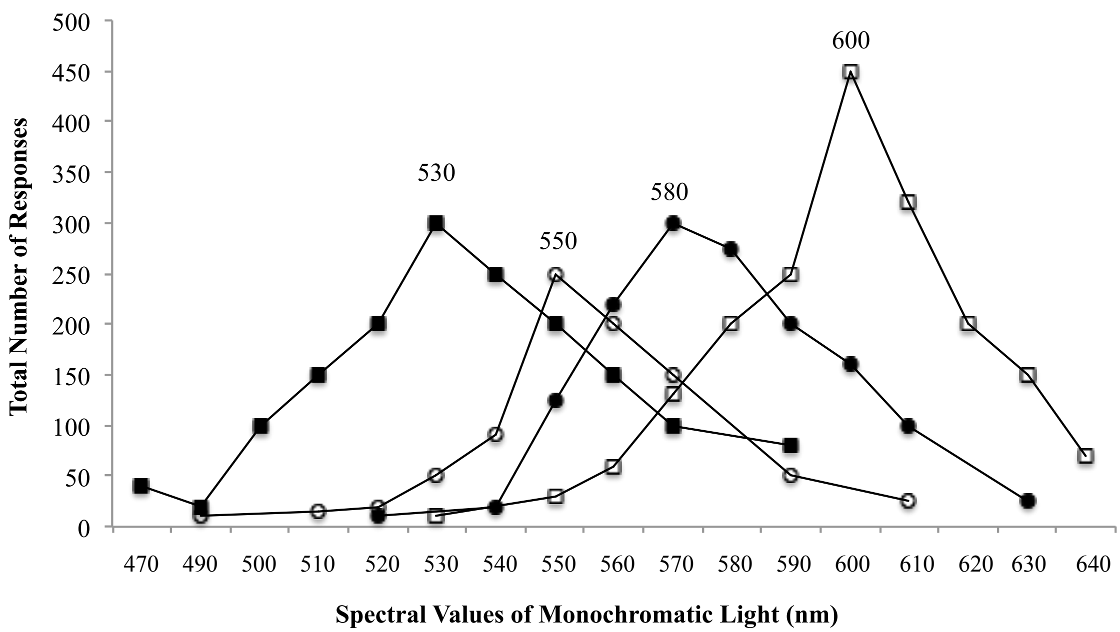


Figure 3. Data from the study by Guttman and Kalish (1956).

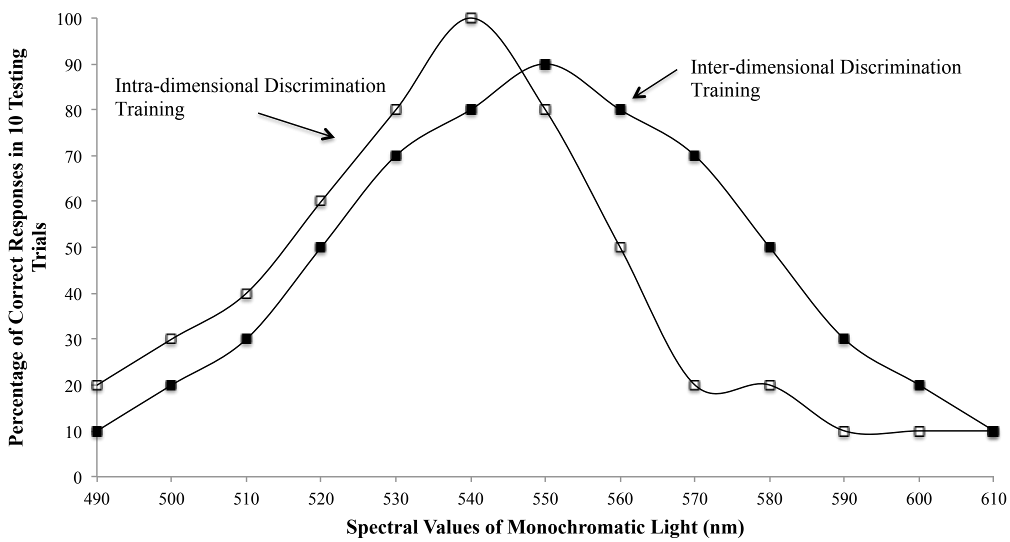


Figure 4. A hypothetical graph for rearrangement gradients along the stimulus dimension of monochromatic light, subsequent to inter-dimensional discrimination intervention (S+: 550 nm; S-: no light), and intra-dimensional discrimination intervention (S+: 550 nm; S-: 570 nm), respectively.

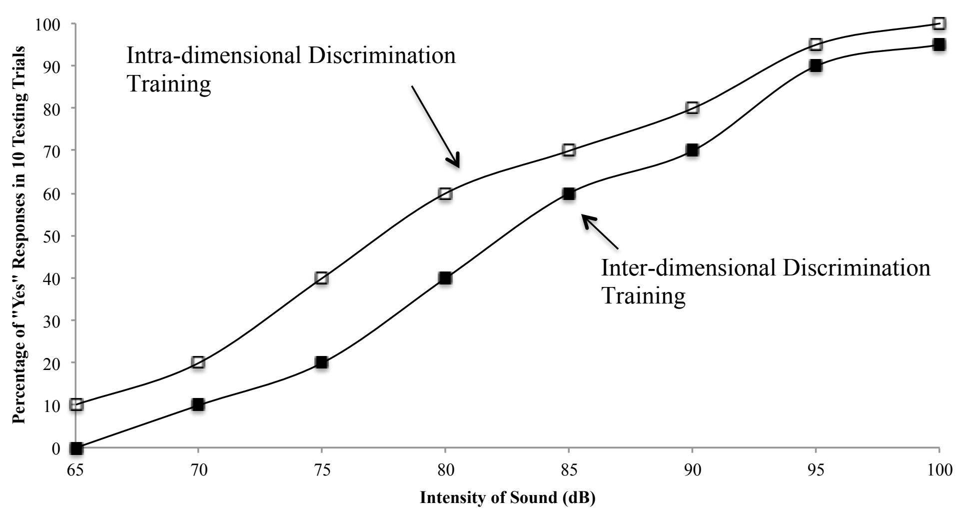


Figure 5. A hypothetical graph for intensity gradients along the stimulus dimension of sound intensity after inter-dimensional discrimination intervention (S+: 85 dB; S-: no sound) compared to intra-dimensional discrimination intervention (S+: 85 dB; S-: 90 dB).

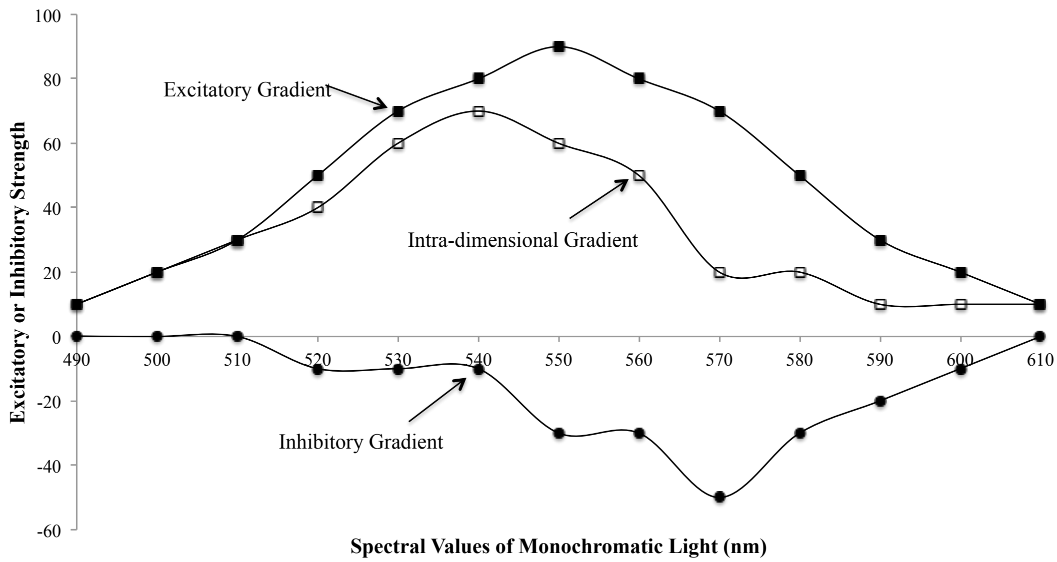


Figure 6. A hypothetical graph of the excitatory gradient, inhibitory gradient, and the intra-dimensional gradient derived from the summation of the excitatory and inhibitory gradients (S+: 550 nm; S-: 570 nm).

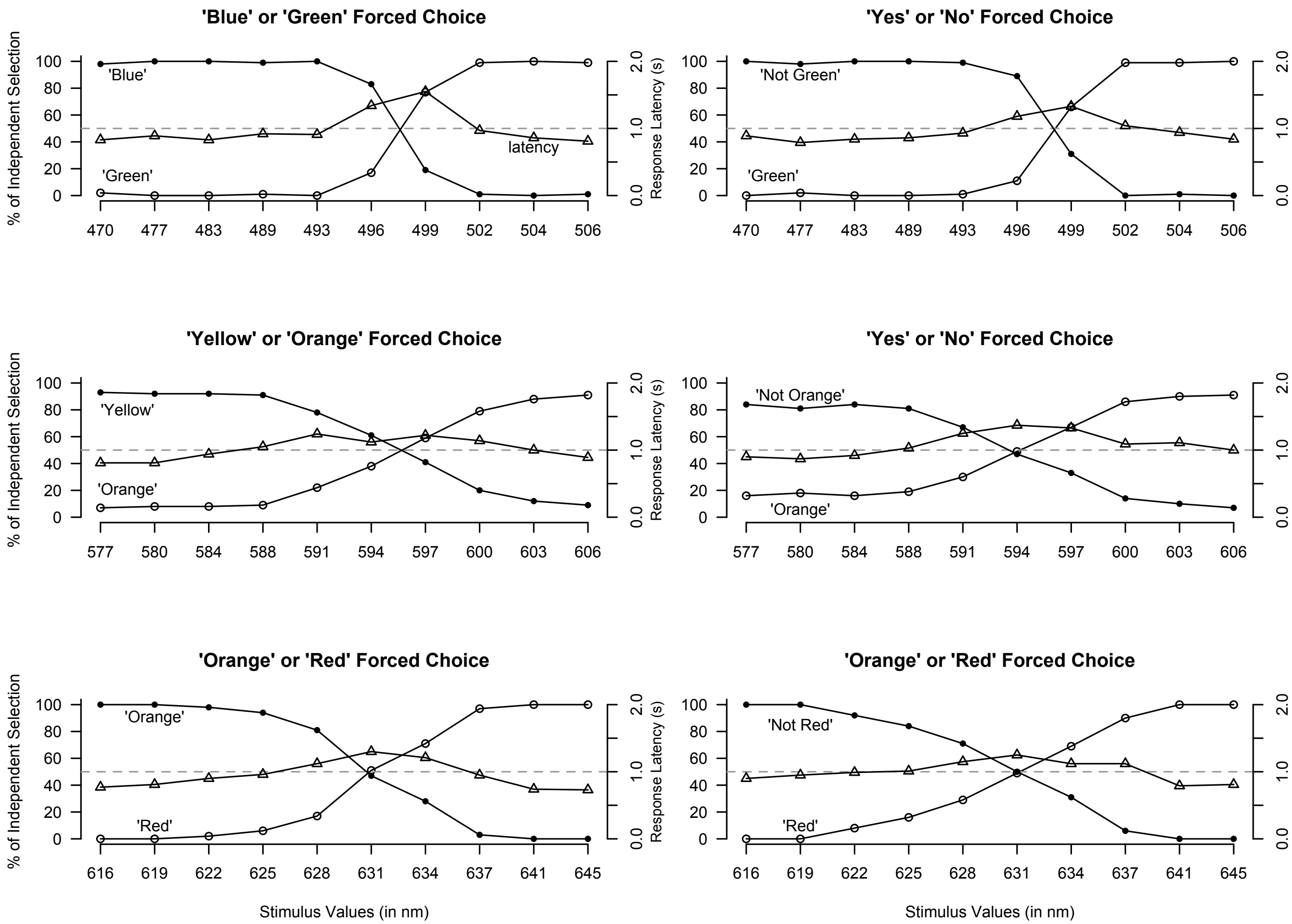


Figure 7. Percent of independent selection for the target choice and response latency for 30 typical adults in each of two response modes across three color sets. (The combined percent was not 100% for all stimuli since “no selection” occurred 31 times out of the total 5,400 opportunities.)

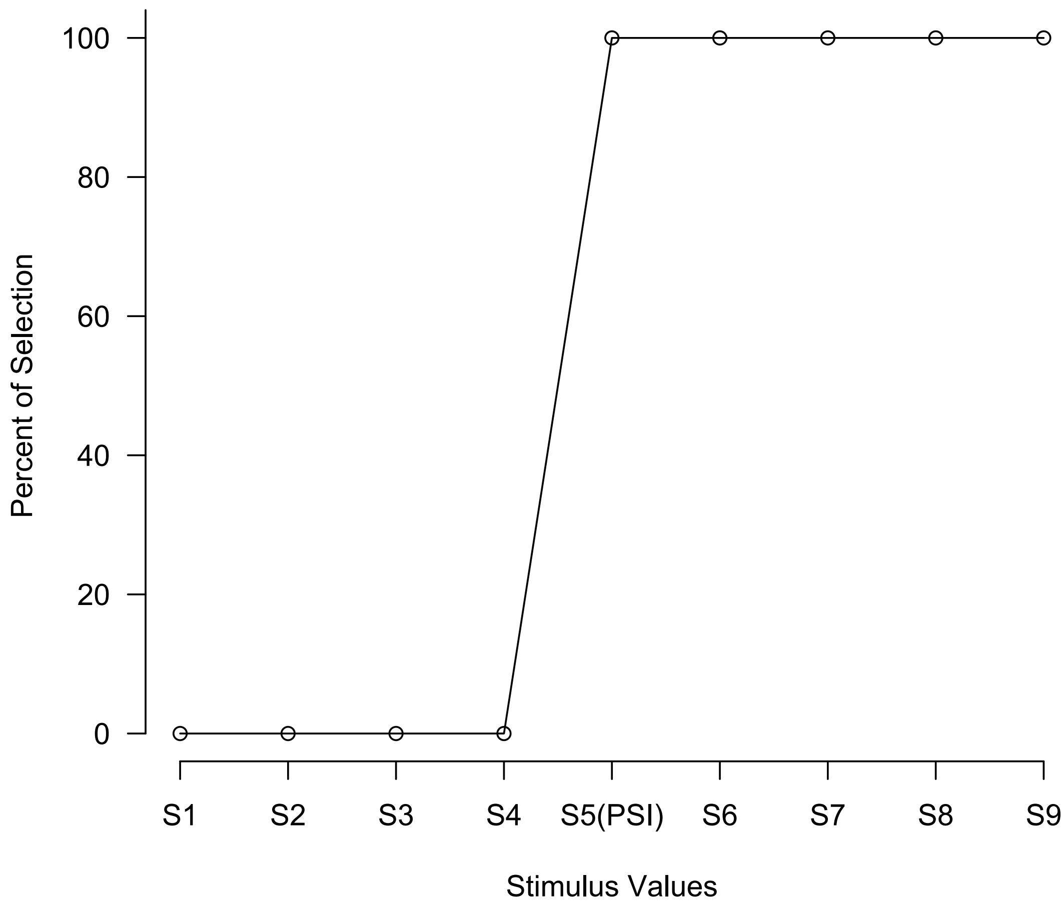


Figure 8. A hypothetical ideal generalization gradient[9].

[9] In this ideal generalization gradient, the stimulus S5 is designated as the PSI stimulus, the stimuli from S6 to S9 are designated as belonging to stimulus Class A, and the stimuli from S1 to S4 are designated as belonging to stimulus Class B. For the stimuli within stimulus Class A, the correct response choice is the same as the one to which the PSI stimulus corresponded. For the stimuli within stimulus Class B, the correct response choice should be the other option.

(a)

(b)

(c)

Figure 9. Photos demonstrating the placement of the "perfect" color choices.

Figure 10. Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for each of the five participants with moderate to severe IDD during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

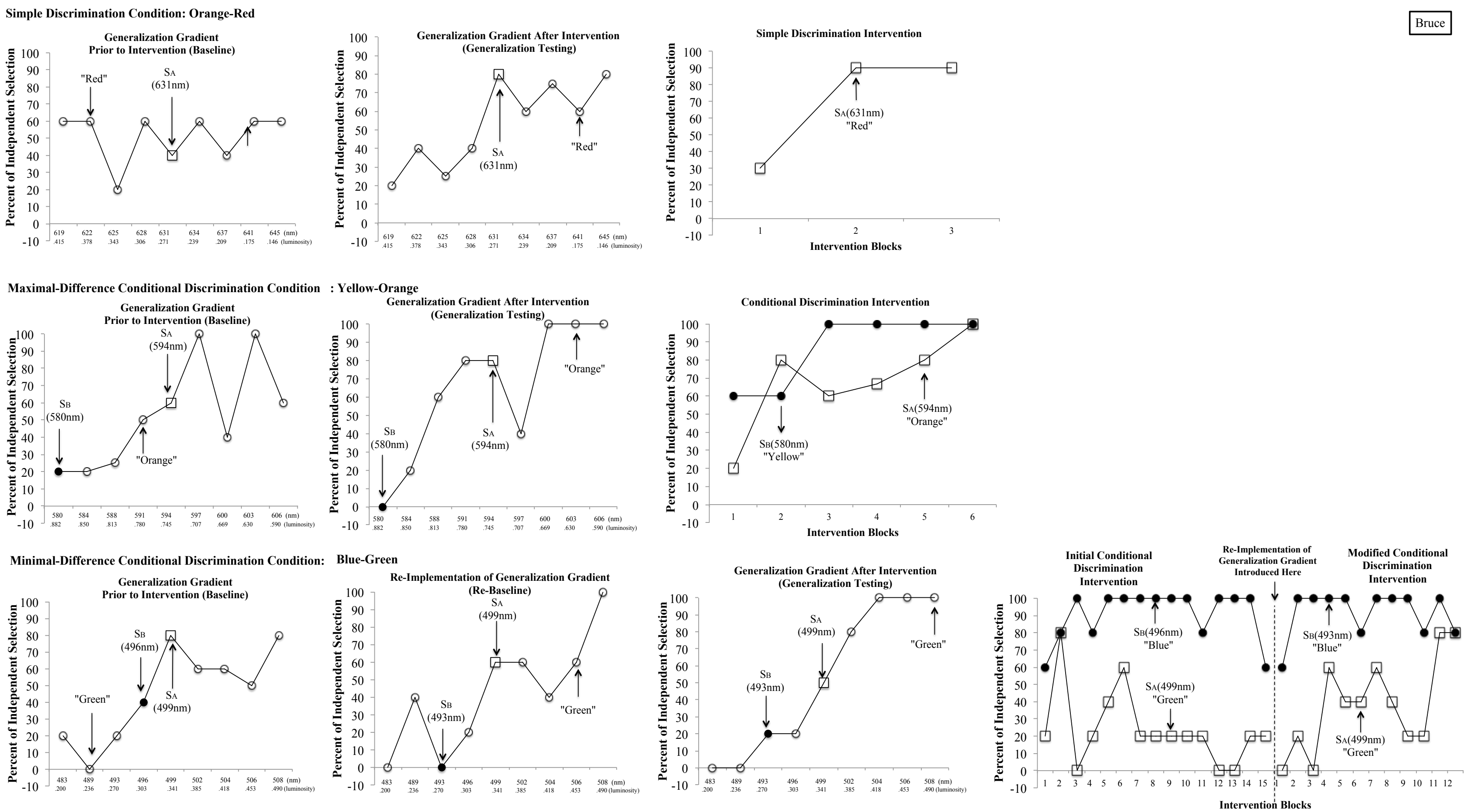


Figure 10(a). Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for Bruce during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

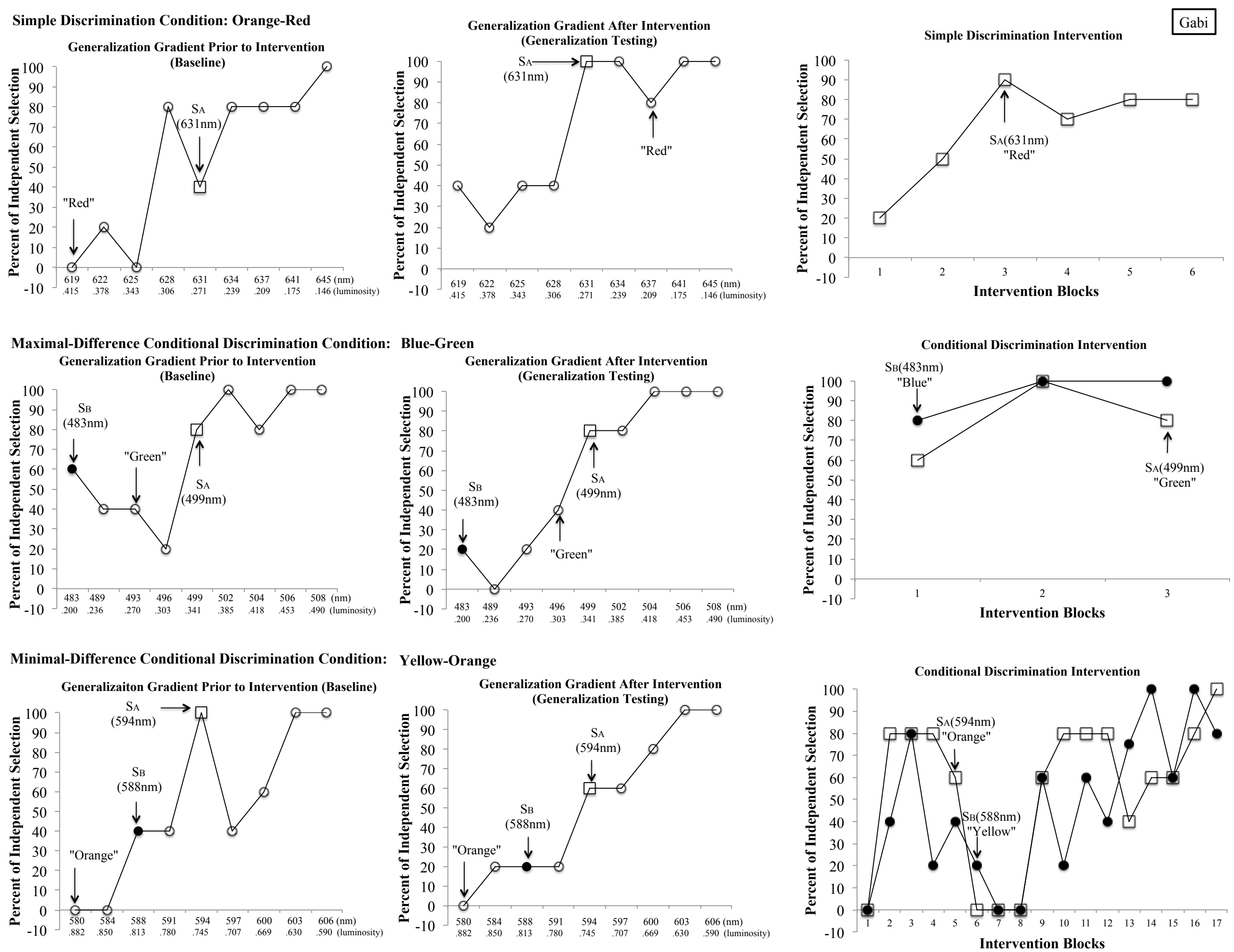


Figure 10(b). Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for Gabi during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

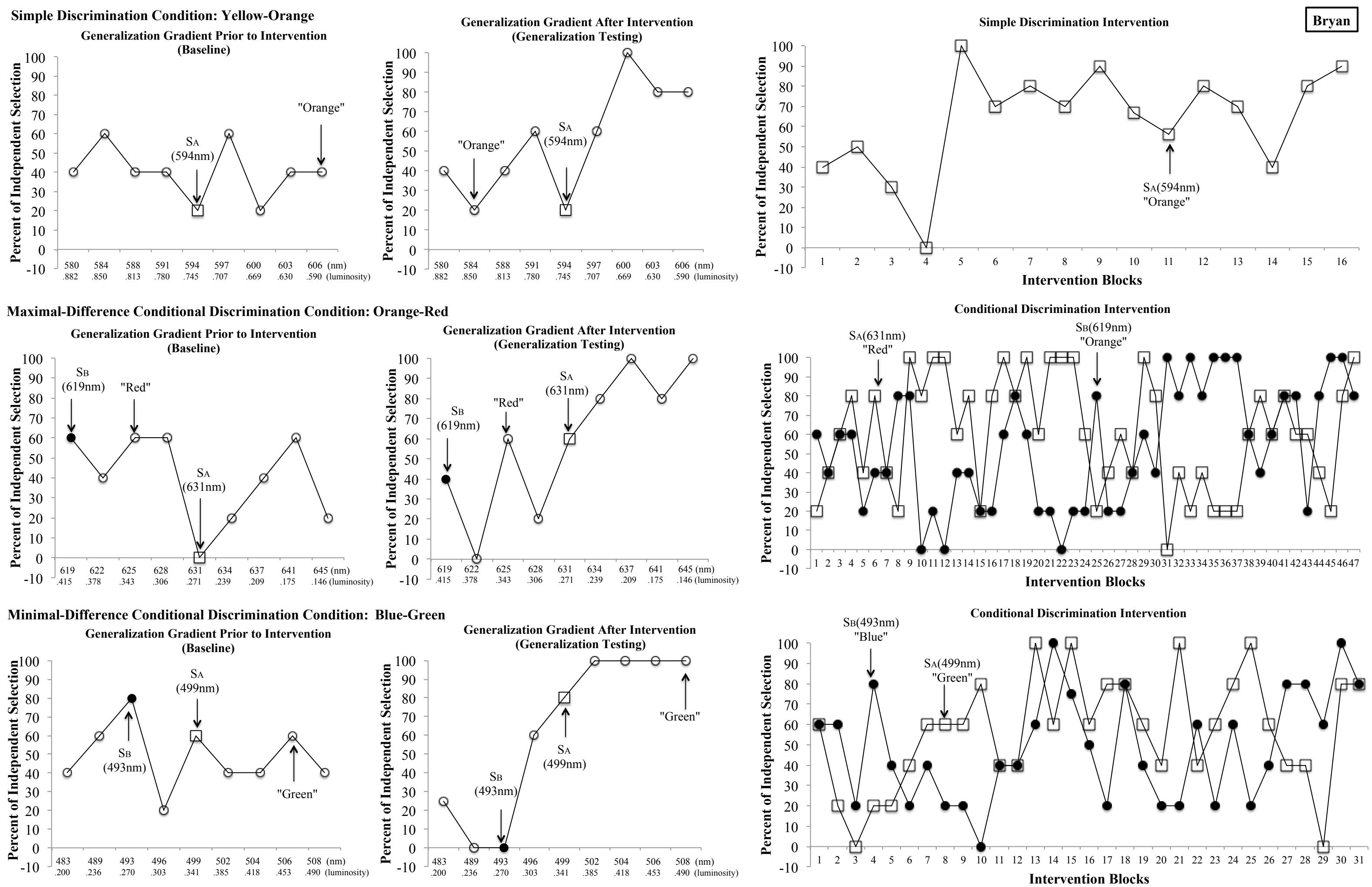


Figure 10(c). Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for Bryan during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

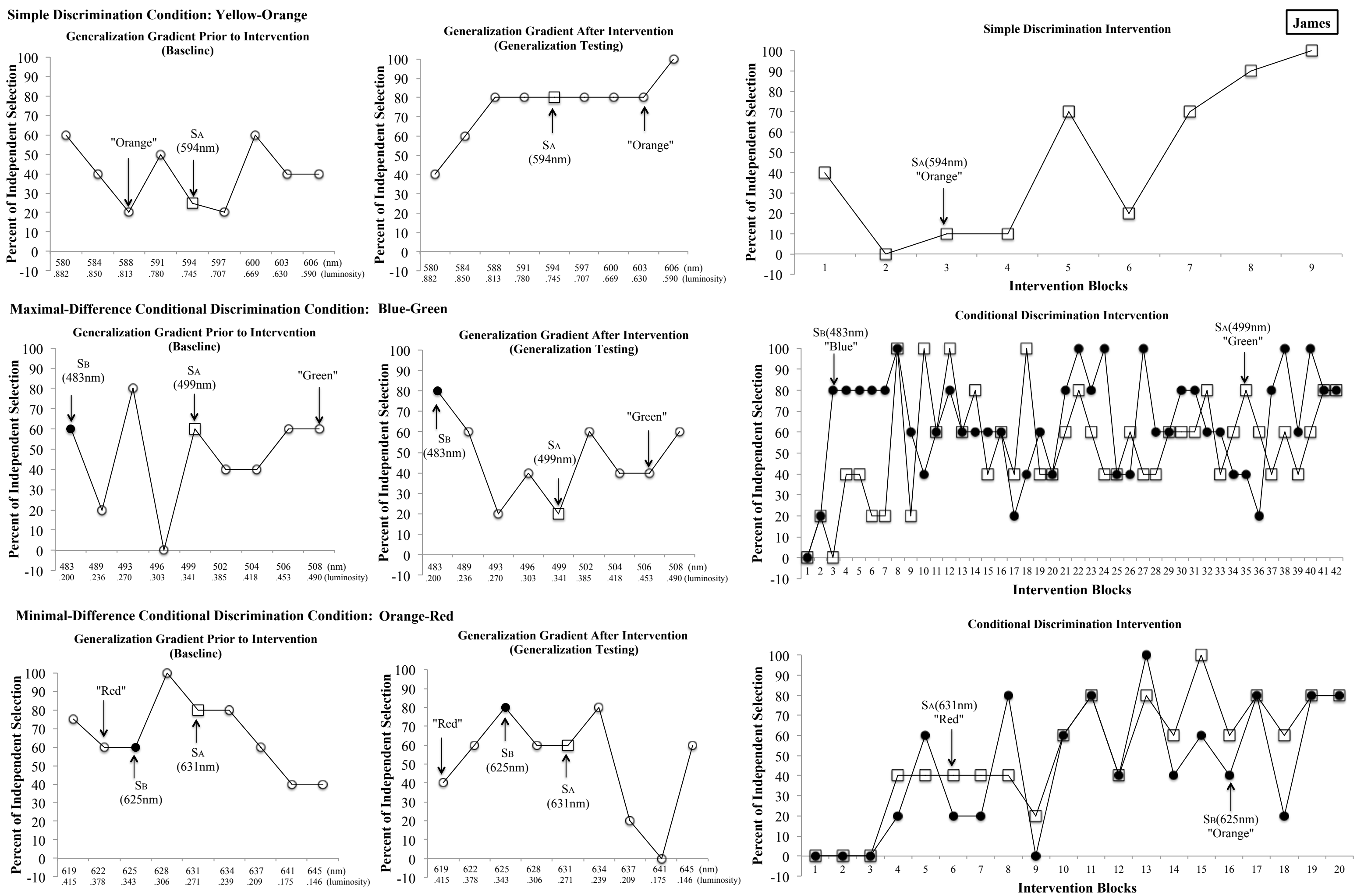


Figure 10(d). Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for James during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

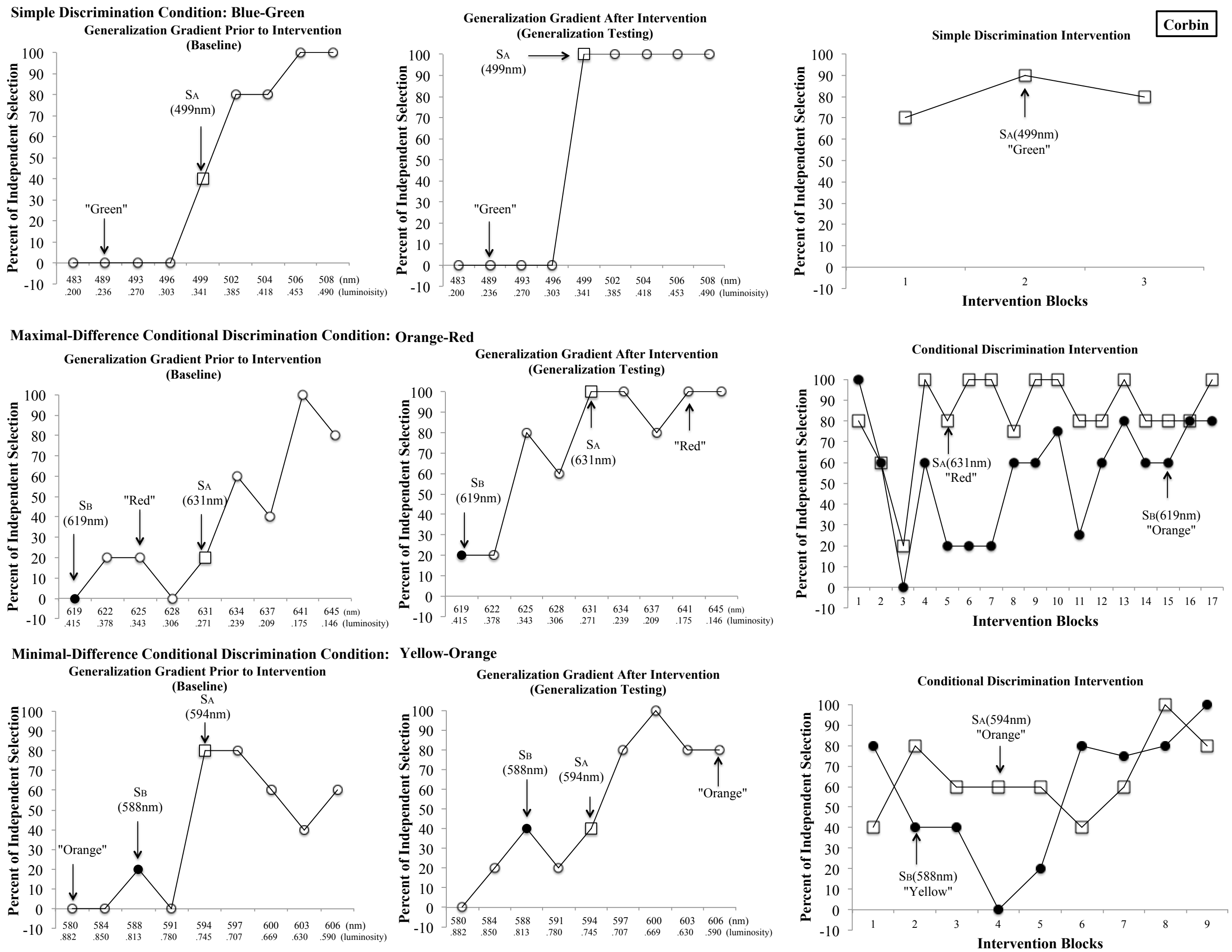


Figure 10(e). Percent of selections for each of nine colors during generalization gradient testing prior to and after intervention and percent correct during blocks of opportunities for Corbin during simple discrimination, maximal-difference conditional discrimination, and minimal-difference conditional discrimination conditions.

Appendix: Method and Results of Pilot Study

Method

All participants first received a color vision screening to make sure they did not have color blindness. Qualified participants subsequently received practice task consisting of two subtasks to get familiar with two response modes (i.e., a forced choice between two specific categories, and a forced choice between "Yes" and "No"). Subsequently, each participant received the color validation experiment consisting of six subtasks. Among the six subtasks, three required the participants to make a forced choice between two target colors at each end of a color set (e.g., "Blue" versus "Green" for a blue-to-green color set). The other three subtasks required the participants to make a forced choice between "Yes" or "No" for each color set (e.g. "Is this 'Green'?")

Participants

Thirty participants included undergraduate and graduate students from the University of Minnesota (18 women, age range: 19-29 years old, median age = 24.5 years) who provided written informed consent in accordance with the University of Minnesota Institutional Review Board (IRB) guidelines. All participants spoke and read English. Twelve participants spoke Chinese as their native language, eight spoke English, two spoke Korean, two spoke Vietnamese, two spoke Spanish, two spoke Turkish, and two spoke Thai. All participants passed a color vision screening (i.e., had demonstrated evidence of no color blindness), and were at least 90% correct on each of two practice subtasks (to be described below).

Setting

Each participant completed all tasks in a conference room containing ambient overhead lighting. The experimenter sat beside each participant during color vision screening and the two practice subtasks. Each participant completed the experiment alone in the room with the experimenter out of the room.

Materials

Laptop-based presentation. A 17-inch Toshiba® laptop was used to present all stimuli on a white background through the Presentation® software (Neurobehavioral Systems, 2004). The screen brightness was set at 50%. For each subtask, a file was created that wrote the programming instructions for Presentation® software. The resulting file was created, allowing keyboard buttons to control stimulus selection by the participant. To execute each subtask, the investigator loaded the corresponding file in the Presentation® software, entered the participant number, and activated the file. The file began with the word "READY?" displayed in white letters on a black background. The presentation of stimuli for each task did not start until the participant pressed "Enter" key on the keyboard. Each participant response was automatically saved into the logfile of the Presentation® software.

Picture stimuli. Two colored photographs of a dog and two colored photographs of a cat obtained from the Google search engine were used during practice opportunities. Three sets of color stimuli (see Table 2) were used in the experiment, with each color stimulus in a shape of circle and in the size of 3.20 inches in diameter. Each set consisted of 10 different color stimuli. One set varied along the blue to green spectrum; a second set varied along the yellow to orange spectrum; and the third set varied along the orange to red spectrum. Thus, a total of 30 different color stimuli were created.

The color differences between any two adjacent color stimuli within each color set were equated by the index of wavelength (with the measurement unit of nanometer [i.e., nm]), concurrently adjusted by the luminosity[10]. In basic research on stimulus control with animals as

---

[10] Luminosity is the measure of the effectiveness of lights of different wavelengths (Sharpe, Stockman, Jagla, & Jägle, 2005). The term was introduced by the International Lighting Commission (Commission Internationale de l'Eclairage or CIE) to reflect the human sensitivity to light of different wavelengths in a typically illuminated environment (Wyszecki & Stiles, 1982). It is based on subjective judgements of which of a pair of different-colored lights is brighter, to describe relative sensitivity to light of different wavelengths. The formula of the luminosity function (or visual sensitivity function) is: $V(\lambda) = \frac{\psi_{555.016}}{\psi_\lambda}$, in which λ refers to the wavelength, and ψ refers to the radiant flux (unit: W). In a typically illuminated

participants, the wavelength was the only index when the experimental stimuli were colors (e.g., Guttman & Kalish, 1956; Hanson, 1959). An index of luminosity was used to equate the difference between color stimuli with respect to the effectiveness of different wavelengths to human eyes, with the index of wavelength still taken into account. Table 3 lists the wavelength and corresponding luminosity value for each color stimulus to be validated. Specifically, the wavelengths of the color stimuli in the blue-to-green color set were 470, 477, 483, 489, 493, 496, 499, 502, 504, 506 nm; the wavelengths of the color stimuli in the yellow-to-orange color set were 577, 580, 584, 588, 591, 594, 597, 600, 603, and 606 nm; and the wavelengths of the orange-to-red color set were 616, 619, 622, 625, 628, 631, 634, 637, 641, and 645 nm.

On average, the color stimuli on the blue-to-green spectrum had a range of 36 nm on wavelength (470-506 nm) and .32 units on the scale of luminosity that ranges from 0 to 1 (.13-.45). The color stimuli on the yellow-to-orange spectrum had a range of 29 nm (577-606 nm) on wavelength and .32 units on luminosity (.91-.59). The color stimuli on the orange-to-red spectrum had a range of 28 nm on wavelength (616-645 nm) and .31 units on luminosity (.45-.14). On average, two adjacent color stimuli differed about 3.44 nm on wavelength and .03 on luminosity. In addition, color stimuli were constructed by converting wavelength (in nm) into red, green, and blue (RGB) values. Table 3 also lists the corresponding RGB value for each color stimulus to be validated.

---

environment, humans are most sensitive to the light with the wavelength of 555.016 nm. For a light of wavelength λ, ψ_λ is the radiant flux required to result in the subjective judgment of brightness as the light of the wavelength of 555.016 nm and with a radiant flux of $\psi_{555.016}$. For instance, the light of 555.016 with a radiant flux of 1 mW and the light of 400 nm with 2.5 W produces a judgment of the same brightness, the luminosity value for the light of 400 nm based on the luminosity function is $V(400\ nm) = \frac{10^{-3}}{2.5} = 0.0004$. Based on the way the luminosity value is calculated, it is dimensionless with no specific unit. It is an interval measure with values ranging from 0 to 1 (Sharpe et al., 2005). The value is close to 1 when the wavelength of the light is 555.016 nm, to which humans are most sensitive. The value decreases when the wavelength of the light is higher or lower than 555 nm. Notably, the visible wavelengths for human beings range from about 390 to 700 nm. The original sensitivity values for each wavelength was found on http://web.archive.org/web/20070314050445/http://www.cvrl.org/database/data/lum/ssvl2e_1.txt

Research Design

A 3 (three sets of color stimuli) × 2 (two response modes) within-participants' factorial design was applied. One within-participant factor, color stimuli, had three sets, with one color set consisting of 10 stimuli ranging from blue to green, one color set consisting of 10 stimuli ranging from yellow to orange, and one color set consisting of 10 stimuli ranging from orange to red. The other within-participant factor, response format, had two forced-choice modes. For one mode the participants responded by making a choice between two color choices (i.e., "Blue" or "Green"; "Yellow" or "Orange"; "Orange" or "Red"). For the other mode, the participants were required to make "Yes" or "No" responses toward three questions (i.e., "Is this green?" "Is this orange?" and "Is this red?"). Consistent with the research design, the experiment consisted of six subtasks to encompass all combinations of the three color sets and the two response modes (i.e., 3 × 2 = 6).

Independent variables. There were two independent variables. One involved the two response modes (i.e., a forced choice between two concrete choice options, and a forced choice between "Yes" and "No"). The other one was the difference among color stimuli in each color set in terms of their wavelength and luminosity.

Dependent variables. There were two dependent variables. The first one was the percent of independent selections of each of the two response choice options involved in a subtask (e.g., "Blue" or "Green," or "Yes" or "No") for a specific stimulus presented on the screen. The second one was the response latency that was the duration between the presentation of a specific stimulus on the screen and the participant's button-pressing selection or first selection if more than one selection were made, as measured in seconds.

Procedures

Color blindness screening. Color blindness screening was conducted for the participants prior to their proceeding to the practice and experimental tasks. Each participant took the *Ishihara Color Vision Test* (Waggoner, 2011), a commercially available color vision test and one of the most frequently used test to screen for color vision blindness for adults (Rodriguez-Carmona, O'Neill-Biba, & Barbur, 2012). Each screening implementation required approximately 5 min. All participants passed the test, demonstrating the evidence of no colorblindness.

Practice. All participants first received practice composed of two subtasks. One subtask required of a forced choice between two specific categories, and the other subtask required of a forced choice between "Yes" and "No." Each was designed to familiarize participants with pressing keyboard buttons required to produce a response and familiarize them with the two types of response modes that were required in the experiment. Throughout the practice with both subtasks, the laptop was placed on the table at a viewing distance of approximately 40 inches in front of the participant with the screen of the laptop positioned vertically to the table surface. The same picture stimuli were utilized in both tasks and consisted of two colored dog photographs and two colored cat photographs.

In the subtask involving a forced choice between dog and cat, prior to the implementation the participants read typewritten instructions which indicated that they were to press the button "C" on the keyboard as soon as they saw a cat on the screen, but press the button "D" as soon as they saw a dog on the screen. The participant was instructed to select "Enter" button on the keyboard to start the subtask. Each of the two dog pictures and each of the two cat pictures were presented randomly during each of three opportunities, resulting in a total of 12 presentation opportunities. The inter-stimulus interval was 5 s resulting in the subtask implementation of no longer than 1 min.

In the subtask involving a forced choice between "Yes" and "No," the procedures were identical to those just described with one exception. Prior to implementation, the participants were provided instructions indicating that they should, as soon as possible, press "Y" if a dog appeared on the screen; but press "N" as soon as possible if a cat appeared on the screen. This subtask was matched for duration with the previous subtask with an inter-stimulus interval of 5 s.

Each participant completed both of the two practice subtasks, with the order counterbalanced across participants. That is, fifteen participants first received the practice subtask involving two specific choices (i.e., “Dog” or “Cat”), and the other fifteen participants first received the practice subtask involving “Yes” or “No” responses.

Color validation experiment. All participants who performed with 90% accuracy or better on each of the two practice subtasks, participated in the experiment. The experiment consisted of six subtasks, three of them requiring the participants to make forced choices between colors at each end of a color set (i.e., “Blue” versus “Green;” “Yellow” versus “Orange;” “Orange” versus “Red), and the other three requiring the participants to make forced choices between “Yes” and “No” toward one of the three questions (i.e., “Is this ‘Green’?” “Is this ‘Orange’?” “Is this ‘Red’?”). As it was during practice; throughout the color validation experiment, the laptop was placed on the table at a viewing distance of approximately 40 inches in front of the participant with the screen of the laptop positioned vertically to the table surface.

In the three subtasks involving a forced-choice between colors, one subtask required the participants to judge whether a color stimulus from the blue-to-green color set was “Blue” or “Green” by pressing the buttons “B” or “G” on the keyboard. A second subtask required the participants to judge whether a color stimulus from the yellow-to-orange color set was “Yellow” or “Orange” by pressing the buttons “Y” or “O” on the keyboard, respectively. Finally, a third subtask required the participants to judge whether a color stimulus from the orange-to-red color set was “Orange” or “Red” by pressing the buttons “O” or “R” on the keyboard, respectively. During the subtask implementation, each of the 10 stimuli from one of the three color sets (see Table 2) was presented randomly during each of three opportunities (a total of 30 opportunities). Prior to each subtask, the participants read the typed instructions. For example, prior to the subtask involving the blue-to-green color set the instructions said, “Press the key ‘B’ if the color is blue; Press the key ‘G’ if the color is green“.

In the three subtasks involving a forced choice between "Yes" and "No", the procedures were the same as those described above for the subtasks involving a forced choice between two color options, with one exception in terms of the typewritten instructions prior to implementation. For example, the instructions for the blue-to-green color set said, "Press the key 'Y' if the color is green; Press the key 'N' if the color is NOT green".

The presentation order of the six subtasks was randomized without replacement across participants. Prior to implementing a subtask, each participant was instructed to select "Enter" button on the keyboard to start each subtask and make responses as soon as possible. During each subtask, the inter-stimulus interval was 5 s resulting in the implementation of each subtask no longer than 2 min 30 s. Participants were allowed to take a break as they needed once they completed one subtask prior to proceeding to the next one. Overall participants completed the experiment consisting of six subtasks in about 20 to 25 min., with the break time taken into account.

Data Analysis Procedures

Dependent measures including independent selection and response latency were extracted from the corresponding logfile saved by the Presentation® software. Whenever two or more selections were recorded during an individual opportunity, the participant's first response was extracted. Whenever a "no selection" was recorded, the corresponding opportunity was considered as invalid (31 times out of 5,400 total opportunities, which occurred among 19 participants). Response latency was calculated for each recorded response as well.

The results were first aggregated for each participant and then for the whole group. For each participant, on each subtask, the average percent of independent selection of each response choice (e.g., "Blue" versus "Green," or "Yes" versus "No") and the mean response latency were computed for each color stimulus, and for each of the two response modes, respectively. For the group of 30 participants, the results of each individual participant were averaged, resulting in a group mean percent of independent selection of each response choice and the group mean response latency for each color stimulus, under each of the two response modes, respectively.

For the first purpose of the pilot study addressing the identification of the PSI stimulus for each color set, the group mean percent of independent selection of the target response choice (i.e., the choice of “Green” for the blue-to-green color set; the choice of “Orange” for the yellow-to-orange color set; and the choice of “Red” for the orange-to-red color set,) as well as the group mean response latency for the 10 stimuli within each color set was scrutinized for both response modes. Within each color set the stimulus, on which two response choices were most equivalently likely (i.e., the group mean percent of independent selection of the target response choice was close to 50%) and the response latency was generally longest across both modes, was determined as the potential PSI stimulus for that specific color set.

To identify the PSI stimulus for each color set in each of the two response mode conditions a paired sample *t*-test was conducted for the group mean percent of independent selections of the target response choices. Additionally, a paired sample *t*-test was conducted for the group mean response latency between two response modes across the 10 stimuli. In addition, for each stimulus within a color set a paired sample *t*-test was conducted for the individual mean percent of independent selection of the target response choice as well as the individual mean response latency between two response modes across 30 participants.

With the PSI stimulus identified for each color set, as an ad-hoc analysis a within-subject two-way analysis of variances (ANOVAs) was conducted to examine the main effect of different color sets, the main effect of different response modes, and the interaction effects of color sets and response modes, upon the performance on the potential PSI stimulus across 30 participants, in terms of both the percent of independent selection of the target choice and the response latency.

Next, with the PSI stimulus identified for each color set, the finalized color set was created for the experiment to be reported. This required the identification of four color stimuli on the left side of the PSI stimulus and four color stimuli on the right side of the PSI stimulus to create a finalized color set consisting of nine color stimuli including the identified PSI stimulus. The identification of nine color stimuli for each color set was within the reasonable range of the number of stimuli used in generalization testing in most of the stimulus generalization literature (e.g., Eckerman, 1970; LaBerge, 1961) which included learners with moderate to severe IDD.

With the determination of the three finalized color sets to be used in the experiment, the group mean of independent selection of the target choice for the nine stimuli within each color set under each response mode was obtained as the generalization gradients for typically developing adults. Subsequently, the residual sum of squares (RSS) was calculated for each generalization gradient by comparing to the ideal generalization gradient.

As Figure 8 shows, in the ideal generalization gradient, the percent of independent selection of a target choice was 100% in the presence of the PSI stimulus and each of the four stimuli that belonged to the designated stimulus Class A. Meanwhile, the percent of independent selection of the same target choice was 0% in the presence of each of the four stimuli that belonged to the designated stimulus Class B. Herein, for the stimuli within the designated stimulus Class A, the correct response choice is the same as the one to which the PSI stimulus corresponded (i.e., the target choice). For the stimuli within the designated stimulus Class B, the correct response choice should be the other option. More specifically, in this study on a specific color set, one specific color choice was designated as the correct response given the PSI stimulus and the stimuli within the stimulus Class A; whereas the other color choice was designated as the correct response given the stimuli within the stimulus Class B. To calculate the RSS value for each actual generalization gradient for each color set under each response mode, the predicted value from the ideal generalization gradient were first subtracted from the observed value from the actual generalization gradient on each corresponding color stimulus, and then the result was squared. Finally the sum of all the squared numbers was the RSS value. For each color set, the RSS values for generalization gradients under two response modes were further averaged to indicate the overall discrepancy between the typically developing adults' performance and the ideal performance with both response modes taken into account.

Results

Table 4 shows the aggregated results of the percent of independent selection for the target response across the 10 color stimuli for each color set. Table 5 shows the aggregated results of the response latency across the 10 color stimuli for each color set. Table 6 and Table 7 show the within-subject two-way ANOVAs results in terms of the PSI performance across 30 participants. Table 8 shows the results of the RSS values for each of the three finalized color sets. Figure 7 shows the line graphs of the percent of independent selection for each target response and response latency under two response modes across three color sets. The results for each color set are described as below.

Blue-to-Green Color Set

Across the 10 color stimuli within the blue-to-green color set (i.e., 470, 477, 483, 489, 493, 496, 499, 502, 504, and 506 nm), the mean percent of independent selections of "Green" was 39.5% when a forced choice between "Blue" and "Green" was required, and 37.8% when a forced choice between "Yes" and "No" was required. Results from a paired sample *t*-test provided the evidence that there was no statistically significant difference in the mean percent of independent selection between the two response modes across the 10 stimuli within the blue-to-green color set ($t[9] = 1.36$, $p = .21$). Also, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. Results shown in Table 4 demonstrated that there was no statistically significant difference in the percent of independent selection between the two response modes for each of the 10 color stimuli.

In terms of response latency, across the 10 color stimuli within the blue-to-green color set, the mean response latency was 0.99 s when a forced choice between "Blue" and "Green" was required, and 0.96 s when a forced choice between "Yes" and "No" was required. Results from a paired sample *t*-test provided the evidence that there was no statistically significant difference in the mean response latency between the two response modes across the 10 stimuli ($t[9] = 0.82$, $p = .43$). Also, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. Results shown in Table 5 revealed that apart from the stimulus of 496 nm, no statistically significant difference was found between the two response modes for all 30 participants on the other nine stimuli. Participants appeared to respond somewhat quicker to the stimulus of 496 nm when they were required to make a forced choice between "Yes" and "No."

Results indicated that across the two response modes participants consistently emitted the most uncertain judgment on the color stimulus of 499 nm. When the participants were required to make a forced choice between "Blue" and "Green," seventy-seven of them considered the color stimulus of 499 nm as "Green" and 19% of them considered it as "Blue." When the participants were required to make a forced choice between "Yes" and "No" toward the question whether the presented color stimulus was green, for the color stimulus of 499 nm 66% of the participants made the judgment of "Yes" and 31% of them made the judgment of "No." In addition, across both response modes, the participants took the longest time to respond to the stimulus of 499 nm, providing further evidence for the uncertainty on the stimulus of 499 nm.

Yellow-to-Orange Color Set

Across the 10 stimuli within the yellow-to-orange color set (i.e., 577, 580, 584, 588, 591, 594, 597, 600, 603, and 606 nm), the mean percent of independent selection of "Orange" was 40.9% when a forced choice between "Yellow" and "Orange" was required, and 48.2% when a forced choice between "Yes" and "No" was required. Results from a paired sample *t*-test provided the evidence that there was statistically significant difference between the two response modes across the 10 stimuli ($t[9] = -6.48$, $p < .001$). Also, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. Results shown in Table 4 demonstrated that on the color stimulus of 588 nm, the percent of independent selection as orange was higher when the participants were required to make a forced choice between “Yes” and “No.” Responses on the other nine color stimuli did not seem to be substantially different between two response modes.

In terms of response latency, across the 10 stimuli within the yellow-to-orange color set, the mean response latency was 1.02 s when a forced choice between “Yellow” and “Orange” was required, and 1.09 s when a forced choice between “Yes” and “No” was required. Results from a paired sample *t*-test provided the evidence that there was statistically significant difference between the two response modes across the 10 stimuli ($t[9] = -2.30$, $p = 0.046$). That is, overall it took longer for the participants to make a forced choice between “Yes” and “No” on the yellow-to-orange color set. Also, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. Results shown in Table 5 showed that there was a statistically significant difference on the color stimulus of 594 nm between these two response modes. That is, it took longer for the participants to make a choice between “Yes” and “No” on the color stimulus of 594 nm compared to making a choice between “Yellow” and “Orange” on this color stimulus.

Overall, there was some variation in the outcomes between the two response modes. When the participants were required to make a forced choice between “Yellow” and “Orange,” the responses to both the color stimuli of 594 nm and 597 nm were close to the 50% chance level. Specifically, on the color stimulus of 594 nm, about 40% of the participants judged it as “Orange.” On the color stimulus of 597 nm, about 60% of the participants judged it as “Orange.” However, when the participants were required to make a forced choice between “Yes” and “No,” only the response to the stimulus of 594 nm was closest to the chance level, with 49% of the participants judged it as “Orange.” Thus, overall across both response modes, the response toward the color stimulus of 594 nm was consistently closest to the 50% chance level as shown in Figure 7. It also took the longest time for the participants to make responses at the stimulus of 594 nm, providing additional evidence for the judgment uncertainty.

Orange-to-Red Color Set

Across the 10 stimuli within the orange-to-red color set (i.e., 616, 619, 622, 625, 628, 631, 634, 637, 641, and 645 nm), the mean percent of independent selection of “Red” was 44.4% when a forced choice between “Orange” and “Red” was required, and 46.1% when a forced choice between “Yes” and “No” was required. Results from a paired sample *t*-test provided the evidence that there was no statistically significant difference between the two response modes across the 10 stimuli ($t[9] = -0.92$, $p = .38$). Also, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. Results shown in Table 4 demonstrated that only on the color stimuli of 625 nm, the percent of independent selection as red was higher when the participants were required to make a forced choice between “Yes” and “No.” There was no sufficient evidence to support that their responses on the other nine color stimuli differed between two response modes.

In terms of response latency, the mean response latency was 0.95 s when a forced choice between “Orange” and “Red” was required, and 1.01 s when a forced choice between “Yes” and “No” was required. Results from a paired sample *t*-test provided the evidence that there was statistically significant difference between the two response modes across the 10 stimuli ($t[9] = -2.32$, $p = .046$). That is, overall it took longer for the participants to make a forced choice between “Yes” and “No” on the orange-to-red color set. Additionally, a paired sample *t*-test was conducted for each color stimulus between the two response modes across 30 participants. As Table 5 shows, there was not sufficient evidence to support that the participants differed in terms of response latency between these two response modes on any of the 10 stimuli.

In general, across both response modes, the participants demonstrated highest judgment uncertainty toward the color stimulus of 631 nm. As Table 4, 5 and Figure 7 show, no matter which response mode was required, around half of the participants judged the color stimulus of 631 nm as red, and the participants consistently took the longest time to make the response at the stimulus of 631 nm.

Summary

Overall the results suggested that for each of the three color sets, there was one color stimulus for which the participants showed most uncertainty between two choices for each of two response modes (i.e., the 7th color stimulus [499 nm] on the blue-to-green color set, the 6th color stimulus [594 nm] on the yellow-to-orange color set, and the 6th color stimulus [631 nm] on the orange-to-red color set; see Table 2). These were determined to be reasonable candidates for the PSI stimulus of each corresponding color set.

Although the tendency of ambiguous responding to these three color stimuli was somewhat more clearly exhibited when a forced choice between "Yes" and "No" was required (especially on the blue-to-green color set); the results from the within-subject two-way ANOVAs (see Tables 6 and 7) indicated that the main effects of response modes on the percent of independent selection of the target choice at these three potential PSI stimuli was not statistically significant ($F[1, 29] = 0.045$, $p = .833$), and the main effects of response modes for response latency with each of three PSI stimuli did not yield a statistically significant difference ($F[1, 29] = 0.006$, $p = .936$). In this sense, there was evidence supporting that participants' performance at the three potential PSI stimuli was generally equivalent across the two response modes.

The within-subject two-way ANOVAs demonstrated that the main effect of different color sets on the percent of independent selection of the target choice at the potential PSI stimuli was statistically significant ($F[2, 58] = 6.168$, $p = .004$), although its main effect on the response latency at the potential PSI stimuli was not statistically significant ($F[2, 58] = 1.793$, $p = .176$). Thus, there was evidence suggesting that participants' percent of independent selection of the target choice at the potential PSI stimulus varied across three color sets. In particular, on the color stimulus of 499 nm for the blue-to-green color set, the percent of selection of "Green" choice was 71.5% averaged across two response modes. This was less uncertain than that the 43.5% of selection of "Orange" on the color stimulus of 594 nm for the yellow-to-orange color set, and the 50% of selection of "Red" on the color stimulus of 631 nm for the orange-to-red color set. In addition, statistically significant interaction effects between response modes and color sets on the percent of independent selection of the target choice at the potential PSI stimulus ($F[2, 58] = 3.446$, $p = .039$), and the response latency ($F[2, 58] = 3.856$, $p = .027$) were found. This finding suggested that the influence of color sets on participants' performance at the potential PSI stimulus could further differ between the two response modes.

Participants had less uncertainty on the color stimulus of 499 nm within the blue-to-green color set compared to the identified stimuli of 594 nm and 631 nm for the other two color sets. However, participants' selection at the color stimulus of 499 nm was much less certain than the other nine stimuli (especially the adjacent ones) within the blue-to-green color set, in comparison to their selection at the color stimuli of 594 nm and 631 nm within the corresponding color sets. Consequently, the three color stimuli (i.e., 499, 594, and 631 nm) were designated at the final PSI stimulus for the blue-to-green, yellow-to-orange, and orange-to-red color sets, respectively.