UNIVERSITY OF NEVADA, RENO

A Comparison of the Effects of Differing Mastery Criteria on

Maintenance of Acquisition Skills

A dissertation submitted in partial fulfillment of the

requirements for the degree Doctor of Philosophy in

Psychology

by

Sarah M. Richling

Dr. W. Larry Williams/Dissertation advisor

August, 2017

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

SARAH RICHLING

Entitled

**A Comparison Of The Effects Of Differing Mastery Criteria On Maintenance Of
Acquisition Skills**

be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

W. Larry Williams, Ph.D, Advisor

James E. Carr, Ph.D, Committee Member

Matthew L. Locey, Ph.D, Committee Member

Linda J. Hayes, Ph.D., Committee Member

MaryAnn Demchak, Ph.D, Committee Member

Stephen L. Rock, Ph.D., Graduate School Representative

David W. Zeh, Ph. D., Dean, Graduate School

August, 2017

Abstract

The demonstration of behavioral acquisition and the maintenance of performance following treatment is fundamental within the fields of behavior analysis and education (Baer, Wolf, & Risley, 1968; Freeland & Noell, 2002; Pererira & Winton, 1991; Stokes & Baer, 1977). The acquisition of skills for individuals with intellectual disabilities and autism has historically focused on the attainment of a certain mastery criterion (Luiselli, Russo, Christian, & Wilczynski, 2008). Love, Carr, Almason, and Ingebor Petursdottir (2009) reported that 62% of respondents functioning as professional supervisors of early and intensive behavioral interventions indicated that a certain percentage of correct trials across multiple sessions are required to determine if mastery has been achieved. However, there is a surprising lack of research evaluating and validating the use of specific mastery criterion rules within clinical practice. The current study (Phase I) conducted a survey of clinical practices of Board Certified Behavior Analysts® (BCBA's®) and Doctoral Board Certified Behavior Analysts® (BCBA-D's®). Survey results indicated that the most commonly reported mastery criterion was 80% accuracy for three consecutive sessions. Based upon these results, the current study conducted an empirical evaluation of the extent to which the adoption of this mastery criterion rule (as well as a 90% accuracy rule) resulted in skill maintenance for 8 individuals diagnosed with intellectual disabilities and autism (Phase III). Results are presented below.

Dedication

This manuscript is dedicated to my family and friends who have provided copious amounts of love and support throughout the entire process. This dissertation is dedicated to my father, for always pushing me to move beyond the perceived boundaries of my circumstances without complaint, lending a patient ear when I did just that, and for always being my biggest fan. This dissertation is dedicated to my mother, for teaching me that work can be play and that life is full of interesting things to discover with child-like wonder. Her unconditional love and support have always been an inspiration to me. This dissertation is dedicated to all my family and friends who have provided me with the platform for living my life as a perpetual student.

Acknowledgements

There are many people to acknowledge for making this dissertation possible. First, many thanks to my advisor Dr. W. Larry Williams for his support and guidance throughout my graduate student career. Thanks also to Dr. James Carr for his valuable contribution and feedback on this project, and to my dissertation committee for their intellectual contributions. I would also like to thank Amy Gonzalez for her support and assistance in making this project come to fruition. Thanks to my research assistants Terralyn Tiffer and Kendall Temple for their dedicated support. I would like to give special thanks to Dr. Holly Seniuk, Dr. Ashley Greenwald, Dr. Molli Luke, Dr. Melissa Nosik, and Dr. Genevieve DeBernardis for being such funny, inspirational, and strong women and for helping make graduate school an unforgettable experience. You all were my family away from home. Finally, a general thank you to all the students and faculty in the program for creating a unique and amazing learning environment.

Table of Contents

List of Tables

List of Figures

Introduction

Common Clinical Practices Related to Skill Acquisition and Maintenance

The demonstration of behavioral acquisition and the maintenance of performance following treatment is fundamental within the fields of behavior analysis and education (Baer, Wolf, & Risley, 1968; Freeland & Noell, 2002; Pererira & Winton, 1991; Stokes & Baer, 1977). The acquisition of skills for individuals with intellectual disabilities and autism has historically focused on the attainment of a certain mastery criterion (Luiselli, Russo, Christian, & Wilczynski, 2008). Within the field of education, an Individualized Education Program (IEP) is utilized to determine if a skill has been acquired in the form of an annual goal. These goals are written by assigning an arbitrary mastery criterion to demonstrate that a skill has been learned to a sufficient degree to warrant continuing on to teaching a more complex skill. Luiselli et al. (2008) state that, in general, clinicians have utilized a percentage of correct responding as a measure of skill acquisition. Several non-experimental sources claim that an 80% or 90% accuracy performance is required to demonstrate mastery (e.g., Anderson, Taras, & O'Malley Cannon, 1996; Leaf & McEachin, 1999). Luiselli et al. indicate that a degree of consistency of this level of accuracy must also be achieved in the form of maintaining these levels of responding across multiple sessions, days, or trainers.

Inherent in these practices, is the assumption that training a certain skill to mastery will result in maintenance of that skill (either when utilized alone or in combination with other strategies for promoting maintenance). Response maintenance has been defined as the degree to which a behavior persists over time when all or part of the intervention variables responsible for training the behavior are no longer present

(Freeland & Noell, 2002; Stokes & Baer, 1977). Smith (1999) conducted an analysis of

12 peer-reviewed outcome studies since 1980 across various fields which focused on

early intervention treatments for children with autism. The author reported that only one

of these studies (i.e., McEachin, Smith, & Lovaas, 1993) provided data related to

maintained performance following the termination of treatment. Smith warns that this is

a detrimental omission because skill acquisition during treatment does not ensure

maintenance of responding when treatment is withdrawn.

A survey of clinical practices conducted by Love, Carr, Almason, and Ingebor

Petursdottir (2009) confirmed some of the aforementioned conjectures regarding the use

of mastery criteria as a measure of skill acquisition. The authors reported that 62% of

respondents functioning as professional supervisors of early and intensive behavioral

interventions indicated that a certain percentage of correct trials across multiple sessions

are required to determine if mastery has been achieved. The authors did not provide

additional results related to common mastery criteria practices such as the specific

accuracy percentages utilized and the number of sessions across which these percentages

are applied. In addition, 98% of respondents reported the use of procedures designed to

promote skill maintenance. The only specific strategy reported was the use of

interspersing previously mastered skills with acquisition programs. This strategy has

only been empirically evaluated for its effectiveness at enhancing skill acquisition (e.g.,

Dunlap, 1984). However, Rapp and Gunby (2016) state that there is insufficient support

for this strategy facilitating more rapid response acquisition. More importantly, these

authors reported that maintenance results from studies evaluating task interspersal are

limited because the studies were not designed to answer that particular question and each

study utilized differing mastery criteria. Thus, the only strategy reported to be utilized by participants (Love et al., 2009) to promote maintenance was that of task interspersal, a strategy that has yet to be empirically validated for that purpose. Stokes and Baer (1977) reviewed 270 ABA articles and stated that the most commonly used strategy to promote maintenance (67%) was that of *Train and Hope*. This strategy involves training until a skill is considered mastered and periodically probing for performance later. While any potential maintenance data would be documented, it would not be pursued with further programming. While practitioners may report that their clinical programs involve measures to promote maintenance, these strategies are either limited or based upon insufficient empirical support. As such, one of the most common strategies adopted during skill acquisition is the use of a certain mastery criterion as a measure of skill acquisition and an assumption that this will result in maintained responding.

Thirty-six percent of the participants involved in the Love et al. (2009) survey were master's-level Board Certified Behavior Analysts® (BCBA's®), 28% had earned a master's degree with no certification, 16% had earned a bachelor's degree with no certification, and 12% were bachelor's-level Board Certified Associate Behavior Analysts® (BCABA's®). Only 6% of participants had earned a doctoral degree without certification and 2% with certification (BCBA-D's®). While a survey of common clinical practices among a variety of professionals providing services to individuals with intellectual disabilities is important, it is also important to specifically examine common clinical practices of individuals certified as behavior analysts providing professional services. It would be especially beneficial to investigate clinical practices utilized by

individuals who are potentially responsible for providing behavior analytic supervision as these practices are likely to be adopted by supervisees.

To date, no known research exists involving a similar survey of clinical practices of BCBA's® and BCBA-D's® involved in providing services for individuals with autism and intellectual disabilities. Moreover, the adoption of mastery criterion rules within clinical practice in the field of behavior analysis is reportedly common; however, details regarding specific features of these rules and the extent to which they are utilized have not been reported. As such, a survey of clinical practices of BCBA's® and BCBA-D's® related to mastery criteria seems warranted. One purpose of the current study was to conduct such a survey (Phase I).

Empirical Research on Mastery Criteria and Skill Maintenance

The need for this type of survey is further bolstered by the surprising lack of research on the effectiveness of differing mastery criterion rules on the maintenance of skill acquisition for individuals diagnosed with intellectual disabilities and autism. As evidenced by the literature above, mastery criterion rules have been reported to be common-place in professional service delivery; however, support for the adoption of these rules has not been empirically demonstrated in a systematic fashion.

Most research involving the direct empirical evaluation of differing mastery criteria has involved undergraduate students (e.g., Johnston & O'Neill, 1973; Semb, 1974). Johnston and O'Neill conducted a series of five experimental conditions designed to evaluate the impact of a minimum performance criterion and grade labels on academic performance with 65 college students. Participants in this study participated in performance sessions which involved providing the students with a set of item cards

(related to a specific text or lecture unit) to which the student was required to respond orally. The dependent variable utilized throughout the study was a rate of correct responding per minute. Participants were randomly assigned to one of five conditions. The first condition involved no teacher-defined criterion and students were only told to do their best and that they would be graded on a curve. Three additional conditions involved sequential variations of three levels of correct responses per unit of time. The high criteria were 3.8 items per minute correct, the medium criteria were 3.1, and the low criteria were 2.5. If the reading and responding to items is calculated independently of how much time is taken, the approximate percentages of accuracy across the three levels were 90%, 75%, and 60% correct, respectively. The main conclusion of this study (Johnston & O'Neill, 1973) was that without a set criterion in place, student performance was low and variable and that teacher-defined criteria may influence student performance. No skill maintenance evaluations were conducted.

Semb (1974) evaluated the effects of differing mastery criteria and assignment length during a personalized instruction course with 89 undergraduate students. Participants were randomly assigned to one of two groups. The study utilized a within-group reversal design and three different manipulations of criteria and assignment lengths were evaluated. A high criterion-short assignment condition required students to take content and review quizzes until a mastery criterion of 100% correct was achieved. A low criterion-short assignment condition required students to take content and review quizzes until a mastery criterion of 60% or higher was achieved. A high criterion-long assignment condition required that students take a review quiz until a mastery criterion of 100% was achieved and content quizzes were not required. The dependent variable

utilized during this study was the average percentage correct on first-attempt quizzes across four content quizzes and a review quiz. In addition, performance on five pre-training and post-training achievement tests was evaluated in a multiple-baseline format. Results of this study indicated that a high mastery criterion was more effective than a low mastery criterion. The low mastery criterion resulted in significant decreases in percentages of correct responding on both quizzes and achievement tests. The author states that if an instructor aims to maximize test performance, a high mastery criterion should be utilized. No skill maintenance evaluations were conducted.

Some information regarding mastery criteria may be indirectly derived from research targeting other phenomena. Several studies have been conducted on the use of various data collection methods on skill acquisition and maintenance. Cummings and Carr (2009) evaluated the impact of either a continuous measurement system or a discontinuous measurement system on skill acquisition and maintenance within behavioral treatment programs for children diagnosed with autism. Results of this study suggested that continuous measurement may result in stronger maintenance. The purpose of this study was not to evaluate differing mastery criteria and, as such, a 100% correct for two consecutive sessions mastery criterion was used across experimental conditions. However, data from the continuous measurement condition may indicate that a mastery criterion of 100% correct for two consecutive sessions was typically effective for producing maintained responding at levels between 75% and 100% accuracy at three weeks following acquisition.

Najdowski et al. (2009) conducted a replication of the Cummings and Carr (2009) study and reported that no such differences in maintenance across the two data collection

methods were observed. However, Najdowski et al. utilized an 80% correct responding across two or three sessions mastery criterion for the continuous measurement condition and a 100% correct responding for the first-trial condition. In addition, the authors implemented an interspersed trial phase following mastery. Results indicated that the use of both mastery criterion levels in combination with interspersed trials resulted in an average maintenance of accuracy above 90% across all participants at a three-week follow-up probe session. Individual participant data was not presented.

Finally, one study evaluated the influence of various maintenance-training methods on maintenance for children diagnosed with autism. An unpublished doctoral dissertation was conducted by Weatherly (2008) and involved a comparison of two different maintenance training methods and a control condition across four curricular programs. As the purpose of this study was not to evaluate differing mastery criteria, a 90% accuracy across two consecutive sessions criterion was utilized for all three experimental conditions. In addition, the study focused on maintenance-training methods implemented after acquisition had already been determined by that specific mastery criterion. However, results from the control condition of this study may provide some insight into the effectiveness of a 90% accuracy across two sessions mastery criterion, alone, for producing maintained responding. Although the higher levels of maintained responding during the two overlearning conditions within this study may be attributed to the implementation of the additional variables following acquisition, responding in the control condition always remained above 70%, even during maintenance probes conducted between 40 and 80 days following the initial demonstration of acquisition. These results may suggest that a 90% accuracy across two consecutive sessions may be

sufficient for producing relatively high levels of response maintenance following a significant duration of time. However, several limitations of this study impede the conclusiveness of this interpretation.

Weatherly (2008) reported that a limitation of that study related to the selection of curricular programs to be taught. The author indicated the targeted classroom did not allow for the random assignment of specified curricular programs for experimental purposes. Rather, programs selected were based upon the participants' IEP goals and not based upon a formal assessment of skill deficits. As such, the author suggested that the skills taught may have been previously acquired outside of the investigation. High levels of accurate responding were observed during the first acquisition sessions which suggests participants may have already acquired considerable levels accuracy prior to the investigation. This may have contributed to the high levels of responding observed during maintenance testing and convoluted comparisons among data series.

In addition, the types of curricular programs and particular instructions (discriminative stimuli) utilized by Weatherly (2008) may have resulted in generalization across data series and also contributed to the high levels of maintenance performance across the independent variable data series. Curricular programs included motor imitation, matching letters, matching pictures, and receptive instruction following. The motor imitation, matching letters, and matching pictures curricular programs all involved a visual-identity match-to-sample discrimination. It is possible that mass training of multiple exemplars of this type of discrimination could result in the development of a generalized match-to-sample repertoire such that participants were able to accurately match across the curricular programs, regardless of the specific independent variables

assigned to particular sets. Further, the instructions (discriminative stimuli) utilized for the receptive instruction program (e.g., "tap desk") were identical to those utilized for the motor imitation program (e.g., "tap desk"). Again, the additional training variable implemented for one experimental set could functionally serve as additional training for untargeted experimental sets.

Given the lack of research specifically designed to systematically evaluate differing mastery criteria utilized by behavior analytic practitioners, another purpose of the current study was to evaluate the extent to which specific mastery criterion rules influence response maintenance (Phase III). The current study aimed to circumvent some of the methodological limitations encountered by Weatherly (2008). Specifically, the current study utilized a reliable assessment of skill deficits and discrimination abilities to help determine appropriate skills to be taught to each learner (Phase II). The curricular programs selected did not include skills currently being taught within the typical classroom. In addition, the specific types of skills selected were done so with the consideration of guarding against generalization across experimental data series and curricular programs. Finally, baseline sessions were conducted to ensure the specific skills selected for each individual are not performed at considerable levels of accuracy prior to the onset of skill acquisition training (Phase III).

To reiterate, the purpose of the current investigation was three-fold. The current investigation involved a survey of clinical practices related to the use of specific mastery criterion rules for BCBA's® and BCBA-D's® who provide services to individuals with intellectual disabilities and autism (Phase I). The current investigation utilized the results of this survey to develop an empirical evaluation of the most-commonly reported mastery

criterion as it relates to skill maintenance (Phase III). To determine appropriate skills to be taught during this phase of the investigation, a reliable assessment of skill deficits and discriminative abilities was conducted for all participants (Phase II).

<div align="center">**Phase I: Survey of Clinical Practices**</div>

Method

      Participants. Participants for this phase of the study were recruited by emailing invitations to participate to approximately 15,677 contacts through the Behavior Analysis Certification Board® (BACB®) listserv. Email invitations were sent to individuals registered through the BACB® listserv as following: (1) residing within the United States; (2) certified as a BCBA-D® or BCBA®; (3) indicating a primary emphasis of work in behavior analysis, behavior therapy, education, or positive behavior support; (4) indicating a primary area of work in intellectual disabilities, autism, special education, or college education; and (5) indicating the primary age group of clients served as children or adolescents. This represented approximately 36% of certificants registered through the BACB® listserv. The initial email invitation (Appendix A) indicated the survey was to be completed by individuals involved in service delivery in the area of intellectual disabilities and autism. One hundred and ninety-nine qualifying responses were received to an online survey supported by Qualtrics® (www.qualitrics.com).

      Materials. Participants were provided with a link to access the online survey within the email invitation. The link was active from April 29th, 2016 through June 31st, 2016. The survey (Appendix B) included a total of 23 multiple-choice and fill-in-the-blank questions that addressed (1) the certificant (e.g., certification status, current provision of applied behavior analytic (ABA) services, primary role in behavior analysis,

and educational background); (2) clients served (e.g., if serving individuals with autism/intellectual disabilities and the ages of clients); and (3) clinical practices related to mastery criteria (e.g., dependent variable utilized, number of sessions applied, and influential source of selection). Prior to the administration of the questions, participants were informed of the purpose of the survey. Additionally, participants were informed that submission indicated consent to participate in research, that there were no known risks, and that any question could be left unanswered at the discretion of the participant. The survey began with three potentially disqualifying multiple-choice questions. Participants were asked if they were currently certified as a BCBA® or a BCBA-D® and those who were not were immediately disqualified. Participants were also asked if they currently provided, oversaw/supervised, or consulted on ABA services and those that indicated a "no" response were immediately disqualified. Finally, participants were asked if their involvement in ABA services was in the area of autism/intellectual disabilities and those that indicated a "no" response were immediately disqualified. These qualifying measures were instilled to further ensure that data was only obtained from individuals who were currently certified and were directly involved in the provision, supervision, or consultation of ABA services in the area of autism/intellectual disabilities and resulted in the above-mentioned 199 qualifying responses. The remainder of the survey took approximately 5 min to complete. Responses were then viewed anonymously via the Qualtrics® website.

Results

The results for each of the areas addressed by the survey are presented below. Note that the percentages reported are based on the number of responses for a specific question which did not necessarily equal the total number (199) of survey respondents.

Certificant and clients served. Five questions addressed general characteristics of the certificant and the clients served. Eighty-four percent ($n = 168$) of respondents indicated that they were currently certified as a BCBA® and 16% ($n = 31$) of respondents indicated that they were currently certified as a BCBA-D®. Seventy-nine percent ($n = 158$) of respondents indicated that the highest degree held in behavior analysis was a master's degree and 21% ($n = 41$) of respondents indicated that the highest degree held in behavior analysis was a doctoral degree. With respect to the year in which a degree in behavior analysis was earned, responses ranged from the years of 1985 to 2016 ($n = 149$). Sixty-six percent ($n = 98$) of respondents indicated they received their degree between the years of 2007 and 2016, 18% ($n = 37$) between the years of 1997 and 2006, and 9% ($n = 14$) between the years of 1985 and 1994. The highest percentage (74%) of respondents ($n = 147$) indicated their primary role in behavior analysis was that of a practitioner, 16% ($n = 32$) indicated their primary role was that of an administrator, 9% ($n = 18$) indicated their primary role was that of a faculty member, and 1% ($n = 1$) indicated their primary role was that of a student. Regarding the age of the clients served, results showed that an average of 32% clients served fell between the ages of 0-6 years of age, 28% fell between the ages of 7-12 years of age, 12% fell between the ages of 13-17 years of age, and 5% were 18 years and older.

Clinical practices related to mastery criteria. Sixteen questions addressed clinical practices related to the mastery criteria used within the respondents' clinical

practice. Figure 1 is a flow chart depicting the series of questions to which survey participants were exposed based upon conditional responses. Regarding the basis for the most common mastery criterion used, results indicated that the majority of participants (68%, $n = 132$) utilize a certain percentage of correct trials (i.e., session-based), 28% utilize a certain number of correct trials in a row ($n = 55$), and 4% utilize a rate of responses per unit of time ($n = 7$). For participants that indicated they utilize a certain percentage of correct trials, 57% ($n = 75$) responded that they apply a percentage of correct trials criterion across multiple sessions with additional variables and 35% ($n = 46$) responded that they apply a percentage of correct trials criterion across multiple sessions. Only 8% ($n = 10$) of participants indicated that they apply a percentage of correct trials during one session (i.e., a single trial block). Additional results related to the most common mastery criterion reported by participants (i.e., a certain percentage of correct trials) are presented below. Additional results from respondents indicating the utilization of a certain number of correct trials in a row or a rate of response per unit of time have not been included.

Fifty-two percent of individuals ($n = 24$) indicating that they utilized a certain percentage of correct trials across multiple sessions reported using an 80% criterion across those sessions, 28% ($n = 13$) indicated that they applied a 90% criterion, 11% ($n = 5$) indicated that they applied between an 81% and an 89% criterion across multiple sessions, 7% ($n = 3$) indicated that they applied a 100% criterion across multiple sessions, 2% ($n = 1$) indicated that they applied between a 91% and a 99% criterion across multiple
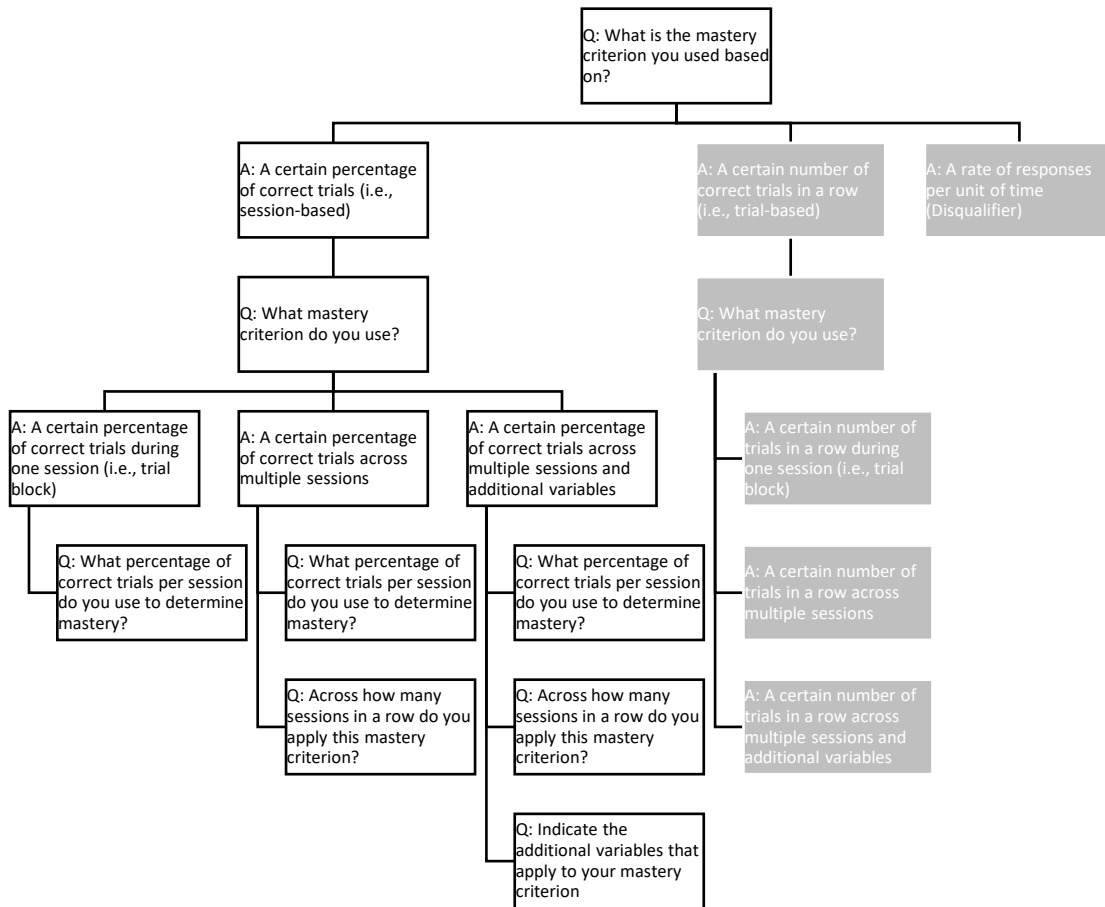
Figure 1. Flow chart depicting the series of questions to which survey participants were exposed based upon conditional responses.

sessions, and 0% indicated that they utilized a below 80% criterion across multiple sessions. Results showed that these percentage correct criteria were applied across three (50%, $n = 23$), 2 (26%, $n = 12$), more than four (20%, $n = 9$), or four (4%, $n = 2$) consecutive sessions.

Fifty-five percent of individuals ($n = 41$) indicating that they utilized a certain percentage of correct trials across multiple session with additional variables reported using an 80% criterion across those sessions, 26% ($n =19$) indicated that they applied a 90% criterion across multiple sessions, 8% ($n = 6$) indicated that utilized a 100% criterion across multiple sessions, 5% ($n =4$) reported using between a 91% and a 99% criterion across multiple sessions, 4% ($n = 3$) indicated utilizing between an 81% and an 89% criterion across multiple sessions, and 1% ($n =1$) indicated utilizing below an 80% criterion across multiple sessions. Results showed that these percentage correct criteria were applied across three (57%, $n = 42$), two (16%, $n =16$), more than four (11%, $n = 8$), and four (11%, $n = 8$) consecutive sessions. With respect to the additional variables applied to these percentage correct mastery criteria across multiple sessions, 81% ($n = 59$) responded "across two or more therapists", 62% ($n = 45$) responded "across two or more environments", 56% ($n = 41$) responded "target interspersed with previously mastered targets", 40% ($n = 29$) responded "reinforcement schedule thinned", 37% ($n = 27$) responded "follow-up probe after a certain period of time", 26% ($n = 19$) responded "first session (trial block) of the day", and 12% ($n = 9$) responded "other".

Finally, survey participants were asked about the primary information source contributing to the use of the specific mastery criterion reported for this survey. The highest percentage of respondents (44%, $n = 83$) specified their supervised experience as

the primary source. Twenty percent ($n = 37$) reported that the mastery criterion utilized was determined by employer policies/requirements. Sixteen percent of respondents ($n = 30$) referenced graduate school training, 10% ($n = 19$) referenced continuing education, 9% ($n = 16$) referenced regulatory requirements (e.g., IEP), and 2% ($n = 3$) referenced a particular funding source.

Discussion

Results from this phase of the study confirmed the aforementioned conjectures of various authors (i.e., Anderson et al., 1996; Leaf & McEachin, 1999; Luiselli et al., 2008) that clinicians often utilize a percentage of correct responding as a measure of skill mastery. The findings of this phase of the study also extend previous survey findings of Love et al. (2009) by demonstrating that the highest percentage of BCBA's® and BCBA-D's® providing services to individuals with autism and intellectual disabilities report utilizing 80% correct trials across three sessions as a measure of mastery. In addition, results indicated that the majority of participants have adopted the use of this mastery criterion level from their supervised experience. Finally, results show the largest number of clients served are between 0 and 12 years of age. Given the common use of the 80% for three consecutive sessions mastery criterion with this pool of clients and the lack of research evaluating the impact of mastery criteria on skill maintenance, the following phases of the current study were designed to empirically evaluate the effectiveness of the most commonly reported mastery criterion rule with the most commonly reported set of participants.

Phase II: Assessment of Basic Learning Abilities-Revised (ABLA-R)

In order to determine appropriate skills to be taught, the Assessment of Basic Learning Abilities-Revised (ABLA-R) (DeWiele, L., Martin, G., Martin, T. L., Yu, C. T., & Thomson, K., 2011) was utilized to formally assess discrimination abilities. This assessment was originally developed by Kerr, Meyerson, and Flora (1977) and was later revised (DeWiele et al., 2011). Martin, Thorsteinsson, Yu, Martin, and Vause (2008) state that the predictive validity of this assessment as it relates to the performance of common tasks by children with intellectual disabilities and autism has been supported since its development.

The ABLA-R provides a measure of an individual's ability to learn various types of discriminations. The assessment involves 6 levels and the specific types of discriminative abilities evaluated are: (a) a simple motor response, (b) a positional discrimination, (c) a visual discrimination, (d) a visual quasi-identity match-to-sample discrimination, (e) a visual non-identity match-to-sample discrimination, and (f) an auditory-visual combined discrimination (Williams, 2015).

Method

Participants. Participants for this phase and the following phase (experimental assessment of mastery criteria) were selected from a comprehensive life skills classroom for students with mild to moderate support needs within a public elementary school in northern Nevada. The children selected for participation must have had a diagnosis of intellectual disabilities (ID) or autism and be between 0-12 years of age. Eight individuals participated in this study, all of which had been diagnosed with intellectual disabilities, had normal vision and hearing and were between the ages of 4 and 9.

Oscar was a 5-year-old boy diagnosed with autism and displayed no verbal communication. Lily was a 9-year-old girl diagnosed with autism and displayed limited verbal communication. Britta was a 4-year-old girl diagnosed with intellectual disabilities and Down syndrome and displayed no verbal communication. Harry was a 6-year-old boy diagnosed with intellectual disabilities and Down syndrome and displayed limited verbal communication. Evan was a verbal 6-year-old-boy diagnosed with autism. Sandy was a verbal 7-year-old girl diagnosed with health impairment and intellectual disabilities. Cyril was a verbal 9-year-old boy diagnosed with health impairment and intellectual disabilities. And Adam was a verbal 6-year-old boy diagnosed with health impairment, William's Syndrome, and intellectual disabilities.

Materials. The ABLA-R consists of six different levels requiring stimuli as described by DeWiele et al. (2011). Two large containers were utilized during the assessment and included a cylindrical yellow can with an open top (15.5 cm in diameter and 17.5 cm in height) and a square box with an open top (14cm x 14cm x 10cm) covered in red and white diagonal stripes. In addition, assessment stimuli included a small, solid, yellow cylinder (approximately 8 cm long and 3 cm in diameter), a small, solid cube with red and white stripes (approximately 4 cm x 4 cm x 4 cm), and an irregularly-shaped piece of grey foam (approximately 4 cm in diameter). The revised ABLA assessment also required additional stimuli in the form of printed wooden words "CAN" and "BOX," colored silver and purple, respectively.

Procedures. The ABLA was conducted using the standard ABLA-R testing criteria as described by DeWiele et al. (2011). Data was recorded using paper and pen on correct and incorrect responses throughout the assessment. Prior to conducting the

assessment, a brief multiple-stimulus without-replacement preference assessment (Carr, Nicolson, & Higbee, 2000; Windsor, Piché, & Locke, 1994 was conducted. The first three preferred stimuli selected by each participant were used throughout the assessment. Prior to each level, participants were exposed to a three-step prompting sequence consisting of (a) a demonstration of the task to be assessed, (b) a physically prompted trial, and (c) an opportunity to engage in the required responses independently. Testing trials began immediately following the participant engaging in an independent correct response for each skill within the respective levels.

During testing trials, reinforcement was provided on a continuous schedule in the form of praise and access to tangible items for independent correct responses. Incorrect responses resulted in the implementation of a correction procedure identical to the three-step prompting sequence provided prior to testing. Testing trials continued for each level until eight consecutive correct responses (pass) or eight cumulative errors (fail) were observed. Correct responses that were observed during the three-step prompting sequence were not counted toward the eight consecutive correct responses required to pass a given level. However, incorrect responses during these error-correction procedures did count toward the eight total incorrect responses required for the fail criterion. An incorrect response was defined as the failure to place the assessment object anywhere other than the level-designated container or as holding onto the assessment object for an extended period of time (i.e., 15 seconds).

ABLA-R Level 1, Simple Motor Response. During this level of the ABLA, the participant was required to place the irregularly-shaped piece of foam into the yellow can

for four consecutive trials and into the red and white box for four consecutive trials. A

modeled prompt was provided prior to each testing trial.

ABLA-R Level 2, Position Discrimination. During this level of the ABLA, the

red box and the yellow can were placed in fixed positions in front of the participant

across trials. The participant was required to place the irregularly-shaped piece of foam

into the yellow can which was in a constantly-fixed position on the left.

ABLA-R Level 3, Visual Discrimination. During this level of the ABLA, the red

box and the yellow can were placed in alternating positions in front of the participant

across trials. The participant was required to place the irregularly-shaped piece of foam

in the yellow can.

ABLA-R Level 4, Visual Identity Match-to-Sample Discrimination. During this

level of the ALBA, the red box and the yellow can were placed in alternating positions in

front of the participant across trials. The participant was required to place the small

yellow cylinder inside of the yellow can and place the small cube into the red box.

ABLA-R Level 5, Visual Non-identity Match-to-Sample Discrimination. During

this level of the ABLA, the red box and the yellow can were placed in alternating

positions in front of the participant across trials. The participant was required to place the

wooden word "CAN" inside the yellow can and to place the wooden word "BOX" into

the red box.

ABLA-R Level 6, Auditory-Visual Combined Discrimination. During this level

of the ABLA, the red box and the yellow can were placed in alternating positions in front

of the participant across trials. The participant was required to place the irregularly-

shaped piece of foam in the yellow can when the experimenter said, "yellow can" in a

low-pitched, slow cadence and to place the same foam in the red box when the

experimenter said "red box" in a high-pitched, rapid cadence.

Results and Discussion

Table 1 displays the participants' ABLA-R scores. Oscar and Lily passed three

levels of the ABLA-R, receiving a score of Level 3. Harry and Britta passed four levels,

receiving an ABLA-R score of Level 4. Evan, Sandy, Cyril, and Adam passed all six

levels of the ABLA-R, receiving a score of Level 6. Scores corresponded with anecdotal

observations of each participants' communication skills.

Table 1. Participant Diagnosis, Anecdotal Communication Abilities, and ABLA-R Score				
Participant	*Age*	*Diagnosis*	*Communication*	*ABLA-R Score*
Oscar	5	Autism	Non-Verbal	3
Lily	9	Autism	Limited Verbal	3
Harry	6	Down Syndrome/Intellectual Disabilities	Limited Verbal	4
Britta	4	Down Syndrome/Intellectual Disabilities	Non-Verbal	4
Evan	6	Autism	Verbal	6
Sandy	7	Intellectual Disabilities/Health Impaired	Verbal	6
Cyril	9	Intellectual Disabilities/Health Impaired	Verbal	6
Adam	6	Intellectual Disabilities/William's Syndrome	Verbal	6

The results of this phase of the experiment were utilized when determining

appropriate skills to be taught during the final phase of this study. Specifically, the

curricular programs taught (described later) involved discriminations that fall either at or

below the discriminative ability level as determined by this assessment.

Phase III: Experimental Assessment of Mastery Criteria

Each participant from the previous ABLA-R phase participated in the

experimental phase of the study. This phase was comprised of four experiments.

Experiment 1 was conducted with individuals who scored below ABLA-R level 6.

Experiments 2, 3, and 4 were conducted with individuals who passed ABLA-R level 6.

Experiment 1

Method.

Participants and setting. Four individuals who scored below ABLA-R level 6 (Harry, Oscar, Lilly, Britta) participated in Experiment 1. Sessions were conducted in a 3.5 m x 4.6 m designated therapy room at the school from which they were recruited during normal school operating hours. The experimental room contained a table, chairs for the participant and the experimenter, and materials relevant to the study. In addition, the room contained a shower and a toilet.

Design and procedures. This experiment was designed to evaluate the maintenance of skills following teaching until a mastery criterion of 80% correct responding for three consecutive sessions was met as compared to both a mastery criterion of 100% correct responding for three consecutive sessions and a mastery criterion of 60% correct responding for three consecutive sessions. A multiple-baseline design across participants combined with an alternating treatments design was used to compare the three mastery criteria such that each of the participants could serve as his or her own control, thus reducing inter-subject variability (Martella, Nelson, & Marchand-Martella, 1999). Each of the four participants who scored below a level 6 on the ABLA were trained with three different sets of discriminative stimuli, one set randomly assigned to each of the three conditions (i.e., 100%, 80%, and 60% mastery criteria levels) for one category of skills (described later).

This experiment involved three conditions (i.e., baseline, acquisition, and maintenance probes). Across these conditions, each experimental training session consisted of 10 discrete trials with one of three sets of stimuli. Across all three phases,

each participant rotated among the various sets of experimental training stimuli. For each participant, three to five experimental sessions were conducted per day, three to five days per week.

Data were collected on the percentage of discrete trials in which the participant made a correct response for each 10-trial session (Appendix C). Responses were only recorded for trials during which a child made an actual response to the instruction within the relevant response class. For example, a correct or incorrect response was only recorded if the child pointed to one of the three comparison cards in the receptive identification program and did so without prompting. If a response was not made within the amount of time specified (i.e., 3 seconds), the trial was repeated until 10 actual responses were made. The purpose of this protocol was to ensure that non-responses due to non-compliance were not included in the calculation of the percentage correct and to ensure accuracy of responding was being measured and not simply whether a response occurred (as is common in resistance to extinction literature).

Prior to each 10-trial experimental session, a brief multiple-stimulus without-replacement preference assessment (Carr et al., 2000; Windsor et al., 1994) was conducted. The first three preferred stimuli selected by each participant during these assessments was used for that session. Access to the preferred stimuli were available across all three phases as described below.

Baseline. During baseline, data were collected on the percentage of correct responses during each 10-trial session. No reinforcement was provided for correct responses and no prompting was provided for incorrect responses. Fifteen to 30 second access to preferred stimuli was provided non-contingently on a 1-minute schedule. The

purpose of this protocol was to decrease the probability of non-compliant behavior associated with escape from demands.

Acquisition. Acquisition sessions were identical to baseline sessions except that reinforcement in the form of praise was provided on a continuous schedule for correct responses (i.e., a fixed-ratio 1 schedule) and 15 to 30 second access to the preferred stimuli was provided on a variable-ratio 3 schedule of reinforcement. In addition, incorrect responses resulted in the use of a least-to-most prompting strategy. Prompted responses did not result in reinforcement. Training sessions for a given set of stimuli were conducted until the respective mastery criterion level (i.e., 60%, 80% or 100% for three consecutive sessions) was met. After the relevant mastery criterion level was met for each set, the maintenance probe phase began. Thus, maintenance probes were conducted with some of the sets of stimuli while some of the other sets were still in the skill acquisition phase.

Maintenance probes. The maintenance probe phase was conducted identical to baseline. Maintenance probes were conducted at approximately one week, two weeks, three weeks, and four weeks following acquisition for each stimulus set. Each maintenance probe consisted of a single 10-trial session.

Inter-observer agreement and procedural integrity. A second independent observer independently collected data on inter-observer agreement (IOA) and procedural integrity in-situ or from video recordings of sessions. IOA was assessed on a trial-by-trial basis by dividing the number of trials in agreement by the total number of trials and then multiplying by 100. IOA data was collected for 35% of baseline sessions, 40% of acquisition sessions, and 46% of maintenance probe sessions. For all participants, the

mean agreement for IOA is 100% across baseline, acquisition, and maintenance probe sessions.

The author of the current study conducted experimental sessions approximately 80% of the time, with one of two trained registered behavior technicians™ (RBTs®) (BACB®) conducting the remaining 20% of the sessions. These percentages were fairly equal across each of the children. For procedural integrity, the implementing tutor's behavior was measured during each trial by a second observer to assess whether (a) the instruction (discriminative stimulus) was delivered correctly, (b) the tutor accurately recorded correct or incorrect responses on the part of the participant, and (c) the tutor followed the designated reinforcement schedule or prompting hierarchy (if needed) immediately following the child's response (Appendix D). This procedural integrity score was assessed based on the number of correct behaviors observed divided by the total number of behaviors. Procedural integrity data was collected for 40% of acquisition sessions. Across all participants, procedural integrity scores averaged 100%. Inter-observer agreement for procedural integrity was collected by a third observer and was assessed by comparing the overall procedural integrity score between independent observers for each session. The mean IOA score for procedural integrity data was 100%.

Curricular program.

Receptive identification/visual discrimination. Three sets of two cards were included in this curricular program with each set being randomly assigned to the three mastery criteria levels (i.e., 60%, 80%, or 100%). For each set, only one pre-designated card was assessed or taught throughout all three phases of the experiment. During a given session, two cards from one of the sets were placed on the table in front of the child. The

tutor then provided the appropriate instruction (e.g., "point to heart"). A correct response was recorded if the participant touched the card as instructed by the tutor. The position of the correct card was randomly rotated throughout each 10-trial session (Appendix E).

Results. Responses were graphed and subjected to visual inspection for identification of differential responding among data series (Michael, 1974). Figure 2 depicts the results for the receptive identification/visual discrimination curricular program across participants with an ABLA-R score below Level 6: Harry, Lily, Oscar, and Britta.

For Harry (Figure 2, top panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion series maintained at training levels at week 1. Thereafter, responding decreased to a range of 40% to 60% for weeks 2 to 4. This range is consistent with chance-level responding (i.e., 50% correct given a stimulus array of two). Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series maintained at training levels at week 1. Thereafter, responding decreased to a range of 30% to 60% for weeks 2 to 4. This range is also consistent with chance-level responding. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series decreased to 80% at week 1 and maintained at this level for weeks 2 and 3. At one month following training, responding decreased to 70%. Overall, responding for the 100% for 3 sessions mastery criterion appeared to remain above chance-levels.

For Lily (Figure 2, 2nd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion ranged between 40% and 70% across weeks 1 to
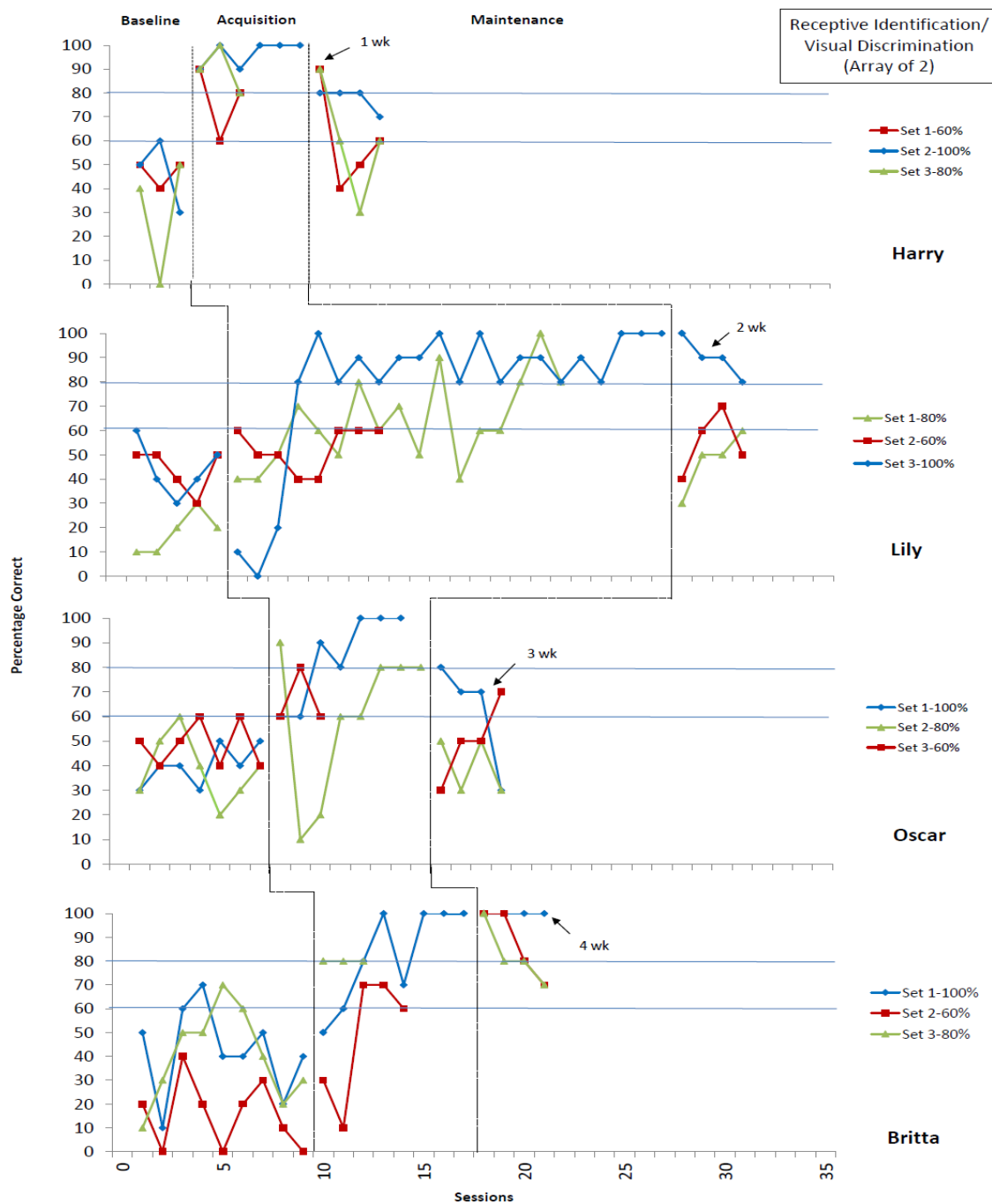
Figure 2. Receptive identification/visual discrimination data across baseline, acquisition and maintenance for Harry, Lily, Oscar, and Britta who scored below ABLA-R Level 6. The figure indicates the percentage of correct responding for each 10-trial session during the 60%, 80%, and 100% for three sessions mastery criteria series.

4. This range is slightly above that of chance-levels. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to 30% at week 1. Thereafter, responding ranged between 50% and 60% correct across weeks 2 to 4. This range is consistent with chance-level responding. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series maintained at 100% at week 1. Responding decreased to 90% for weeks 2 and 3. At one month following training, responding decreased to 80%. Overall, responding for the 100% for 3 sessions mastery criterion appeared to remain above chance-levels.

For Oscar (Figure 2, 3rd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion ranged between 30% and 70% across weeks 1 to 4. This range is consistent with or slightly above that of chance-levels. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to a range of 30% to 50% correct across weeks 1 to 4. This range slightly below chance-level responding and is consistent with baseline levels of responding. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series immediately decreased to 80% at week 1. Responding further decreased to 70% for weeks 2 and 3. At one month following training, responding had decreased to 30%. Overall, responding for the 100% for 3 sessions mastery criterion appeared to remain above chance-levels until one month following the final training session.

For Britta (Figure 2, bottom panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion increased to 100% for weeks 1 and 2.

Responding decreased to 80% at week 3 and 70% at week 4. This range is above that of

chance-levels. Responding during maintenance probes for the 80% correct for 3 sessions

mastery criterion series also increased to 100% at week 1. Responding decreased to 80%

for weeks 2 and 3 and 70% at week 4. This range is above that of chance-level

responding. Finally, responding during maintenance probes for the 100% correct for 3

sessions mastery criterion series maintained at 100% across all four weeks.

Experiment 1 Discussion. Across all participants, 12 maintenance probe data series are

amenable to visual inspection. This includes four maintenance probe data series for each

of the three maintenance criteria levels (i.e., 60%, 80%, and 100% correct responding for

three sessions) (see Appendix H).

For the 60% correct responding for three sessions criterion, two of the four

maintenance probe data series appear to be maintaining at or above criterion levels at

week 1. Across weeks 2 and 4, responding appears to be at or only slightly above

chance-levels for three of the four maintenance probe data series.

For the 80% correct responding for three sessions criterion, two of the four

maintenance probe data series appeared to be maintaining at criterion levels at week 1.

Across weeks 2 and 4, responding appears to have decreased to chance-levels for three of

the four maintenance probe data series. That is, for half of the participants, the 80%

correct responding across three sessions mastery criterion resulted in maintained

responding one week following training. However, responding deteriorated to chance-

levels for three of the four participants after only two weeks following training.

Maintained responding was observed for one participant even after one month (i.e.

Britta). However, responding during maintenance probes had *increased above* the

highest levels observed during training for both the 60% and 80% criteria. One possible explanation for these observed patterns is the participant was receiving relevant training outside of the experimental sessions between training and maintenance probes. However, the participant's teacher stated that no similar stimuli or programs were being utilized in the classroom at any time throughout the entirety of the study.

For the 100% correct responding for three sessions criterion, three of the four maintenance probe data series maintained at 100% at week 1. Across all four maintenance probe data series, responding remained above 70% and up to 100% for at least three weeks. By week 4, responding had only decreased below 70% for one participant.

Overall, data indicate that the 80% correct for three sessions in a row mastery criterion resulted in early maintenance at criterion levels for two of the four participants and had resulted in complete deterioration to chance-levels for the remaining two by week 1. Later maintenance probe data show responding returned to chance-levels for three of four participants by week 2. These data suggest that this criterion, alone, may not be sufficient for maintaining acceptable levels of responding. In fact, across all four comparisons, responding in the 80% for three sessions criterion series was never observed to remain higher than the 60% for three sessions criterion series. Further, these results may tentatively suggest that utilizing a 100% correct responding for 3 sessions criterion is more effective for promoting maintenance of responding.

However, conclusions regarding this data should be drawn tentatively. This experiment only involved four participants who scored relatively low on the ABLA-R. It is possible that higher levels of maintained responding may be observed with individuals

who demonstrate higher performance on discriminative tasks. In addition, this experiment only evaluated performance on one type of task. As such, Experiment 2 was conducted with four additional participants that scored a level 6 on the ABLA-R.

Experiment 2

Experiment 2 served as a direct replication of Experiment 1 except that it included individuals who passed ABLA-R Level 6. In addition, these participants were taught a different curricular skill than in Experiment 1. The discrimative context involved with the selected skill corresponded with participants' performance level during the ABLA-R.

Method.

Participants and setting. Four individuals who passed ABLA-R level 6 (Evan, Cyril, Sandy, and Adam) participated in Experiment 2. The experimental setting was identical to Experiment 1.

Design and procedures. The experimental design and procedures utilized in Experiment 2 were identical to those described in Experiment 1. This involved an evaluation of the maintenance of skills following teaching until a mastery criterion of 80% correct responding for three consecutive sessions was met as compared to both a mastery criterion of 100% correct responding for three consecutive sessions and a mastery criterion of 60% correct responding for three consecutive sessions.

Each of the four participants who passed level 6 on the ABLA-R were trained with three different sets of discriminative stimuli, one set randomly assigned to each of the three conditions (i.e., 100%, 80%, and 60% mastery criteria levels) for one category of skills (a second category of skills was taught to these same participants in Experiment 3).

Inter-observer agreement and procedural integrity. IOA and procedural integrity data were collected as described in Experiment 1. IOA data was collected for 42% of baseline sessions, 37% of acquisition sessions, and 42% of maintenance probe sessions. For all participants, the mean agreement for IOA is 100% across baseline, acquisition, and maintenance probe sessions. Procedural integrity data was collected for 37% of acquisition sessions. Across all participants, procedural integrity scores averaged 100%. The mean IOA score for procedural integrity data was 100%.

Curricular program.

Receptive identification/auditory discrimination. Three sets of three pictures were included in this curricular program with each set being randomly assigned to the three mastery criteria levels (i.e., 60%, 80%, or 100%). For each set, all three cards were assessed or taught throughout all three phases of the experiment. During a given session, three cards from one of the sets were placed on the table in front of the child. The tutor then provided the appropriate instruction (e.g., "point to lemur"). A correct response was recorded if the participant touched the card as instructed by the tutor. The positions and instruction provided were randomly rotated throughout each 10-trial session (Appendix F).

Results. Figure 3 depicts the results for the participants with an ABLA-R score of Level 6 (Evan, Cyril, Sandy and Adam) on the receptive identification/auditory discrimination curricular program.

For Evan, (Figure 3, top panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion immediately decreased to 40% at week 1 and ranged between 20% and 40% across weeks 1 to 4. This range is consistent with chance-
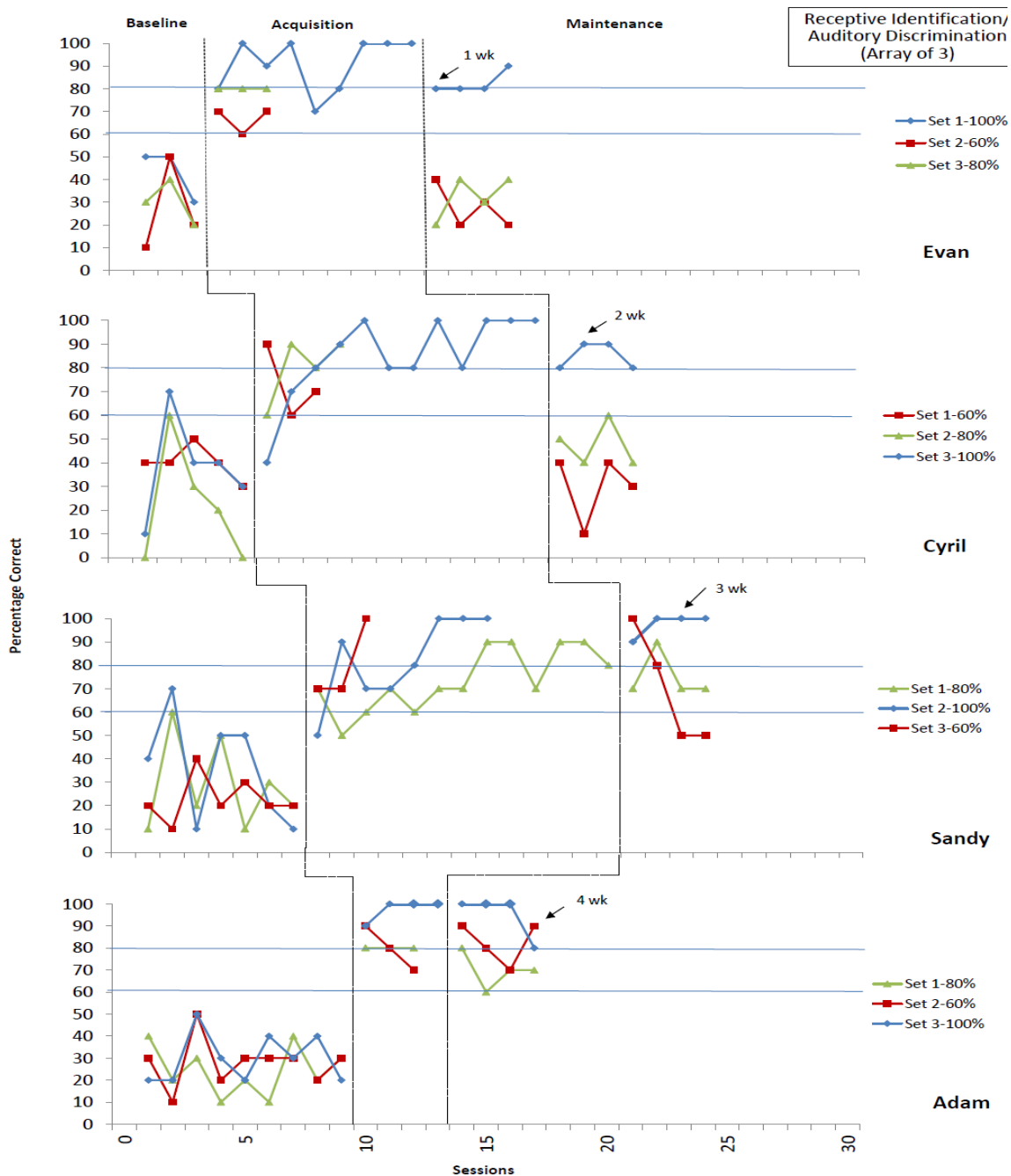
Figure 3. Receptive identification/auditory discrimination data across baseline, acquisition and maintenance for Evan, Cyril, Sandy, and Adam who passed Level 6 of the ABLA-R. The figure indicates the percentage of correct responding for each 10-trial session during the 60%, 80%, and 100% for three sessions mastery criteria series.

level responding (i.e., 33% correct given a stimulus array of three). Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to 20% at week 1 and ranged between 20% to 40% across weeks 1 to 4. Again, this range is consistent with chance-level responding and is comparable to baseline levels of responding. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series immediately decreased to 80% at week 1. Responding during weeks 1 to 4 ranged from 80% to 100% correct.

For Cyril (Figure 3, 2nd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion immediately decreased to 40% at week 1 and ranged between 10% and 40% across weeks 1 to 4. This range is consistent with chance-level and baseline responding. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to 50% at week 1. Correct responding ranged from 40% to 60% correct across weeks 1 to 4. This range is slightly above that of chance-level responding. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series immediately decreased to 80% at week 1. Responding during weeks 2 to 4 ranged from 80% to 90% correct.

For Sandy (Figure 3, 3rd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion remained at 100% at week 1 but showed a decreasing trend within a range of 50% and 80% across weeks 2 to 4. This pattern of responding remained above chance-level and baseline responding. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series slightly decreased to 70% at week 1. Correct responding ranged from 70% to 90% correct across

weeks 1 to 4. This pattern of responding remained above that of chance-level and baseline. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series slightly decreased to 90% at week 1.   However, responding during weeks 2 to 4 was at 100% correct.

For Adam (Figure 3, bottom panel), responding during maintenance probes for the 60% correct for 3 sessions mastery remained at 90% at week 1 and ranged from 70% to 90% across weeks 1 to 4. This pattern of responding remained above chance-level and was comparable to levels of responding observed during training. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series remained at 80% at week 1. Correct responding ranged from 60% to 80% correct across weeks 1 to 4. This pattern of responding remained above that of chance-level and baseline. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series remained at 100% at week 1. Responding continued to occur at 100% until week 4 when responding decreased to 80%.

Experiment 2 Discussion. Across all participants, 12 maintenance probe data series are amenable to visual inspection. This includes four maintenance probe data series for each of the three maintenance criteria levels (i.e., 60%, 80%, and 100% correct responding for three sessions) (see Appendix I).

For the 60% correct responding for three sessions criterion, two of the four maintenance probe data series appear to be maintaining at or above criterion levels at week 1. Across weeks 2 and 4, this level of responding continued except in one instance. For one participant, responding showed a steady declining slope across weeks 2 to 4.

For the 80% correct responding for three sessions criterion, two of the four maintenance probe data series appear to be maintaining at or slightly below criterion levels at week 1. The remaining two data sets show a decrease to chance-levels and baseline level responding. Across weeks 2 and 4, responding remained at chance-levels for those two data sets. The other two data series show responding that continues to occur slightly below criterion-levels. That is, for one of the participants, the 80% correct responding across three sessions mastery criterion resulted in maintained responding one week following training and slightly decreased responding for a second participant. However, responding deteriorated to chance-levels for the remaining two participants at week 1 until week 4. The 80% correct responding across three sessions mastery criterion resulted in responding at or above 70% up to one month following mastery for one participant and at or above 60% for one participant.

For the 100% correct responding for three sessions criterion, three of the four maintenance probe data series maintained at 100% at week 1. Across all 4 maintenance probe data series, responding remained above 70% and up to 100% for at least three weeks. By week 4, responding had only decreased below 70% for one participant.

Overall, data indicate that the 80% correct for three sessions in a row mastery criterion resulted in early maintenance at criterion levels for two of the four participants and resulted in complete deterioration to chance-levels for the remaining two by week 1. These findings are consistent with the results of Experiment 1. For the remaining two participants, the initial and subsequent maintenance probe data show responding at approximately 70%. These data suggest that this criterion, alone, may result in slightly decreased levels of responding for some individuals and result in no maintenance for an

approximately equal numbers of individuals. Similar to Experiment 1, across all four comparisons, responding in the 80% for three sessions criterion series was only observed to remain consistently higher than the 60% for three sessions criterion series in one case. In addition, these results further suggest that utilizing a 100% correct responding for 3 sessions criterion may be more effective for promoting maintenance of responding.

Again, conclusions regarding this data should be drawn tentatively. Although this experiment replicated and extended the results of Experiment 1 with four additional participants, Experiment 2 only evaluated one type of curricular program. Strong conclusions regarding the use of the 80% correct across three sessions mastery criterion should not be made without further replications and evaluations of other types of skills. Experiment 3 was conducted as a first step toward accomplishing this goal.

Experiment 3

Experiment 3 served as a direct replication of Experiment 2. It included the same four individuals who passed ABLA-R Level 6. Experiment 3 only differed from Experiment 2 in that the participants were taught a different curricular skill than in Experiments 1 and 2.

Method.

Participants and setting. Four individuals who passed ABLA-R level 6 (Evan, Cyril, Sandy, and Adam) participated in Experiment 3. The experimental setting was identical to Experiments 1 and 2.

Design and procedures. The experimental design and procedures utilized in Experiment 3 were identical to those described in Experiment 1. This involved an evaluation of the maintenance of skills following teaching until a mastery criterion of

80% correct responding for three consecutive sessions was met as compared to both a mastery criterion of 100% correct responding for three consecutive sessions and a mastery criterion of 60% correct responding for three consecutive sessions.

Each of the four participants who passed level 6 on the ABLA-R were trained with 3 different sets of discriminative stimuli, one set randomly assigned to each of the three conditions (i.e., 100%, 80%, and 60% mastery criteria levels) for one category of skills.

***Inter-observer agreement and procedural integrity*.** IOA and procedural integrity data were collected as described in Experiment 1. IOA data was collected for 42% of baseline sessions, 33% of acquisition sessions, and 48% of maintenance probe sessions. For all participants, the mean agreement for IOA is 100% across baseline, acquisition, and maintenance probe sessions. Procedural integrity data was collected for 33% of acquisition sessions. Across all participants, procedural integrity scores averaged 100%. The mean IOA score for procedural integrity data was 100%.

Curricular program.

Verbal tacting. Three sets of three pictures were included in this curricular program with each set being randomly assigned to the three mastery criteria levels (i.e., 60%, 80%, or 100%). For each set, all three cards were assessed or taught throughout all three phases of the experiment. During a given session, one of the three cards from one of the sets was presented to the child and the tutor said "What is this?". A correct response was recorded if the participant provided the appropriate verbal response assigned to the presented picture. Each phase of the program involved three pictures, presented randomly by the tutor for each 10-trial session (Appendix G).
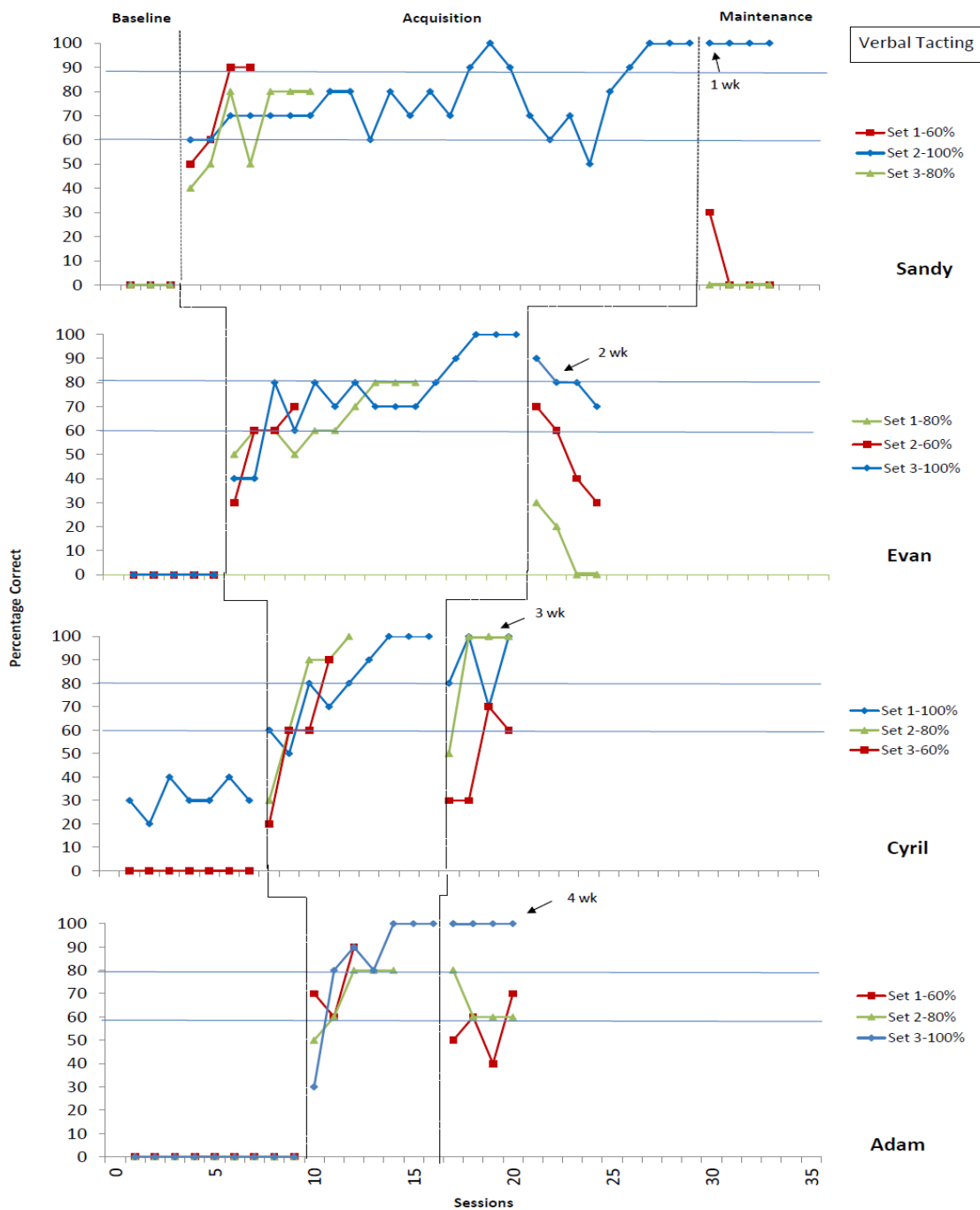
Figure 4. Verbal tacting data across baseline, acquisition and maintenance for Evan,

Cyril, Sandy, and Adam who passed Level 6 of the ABLA-R. The figure indicates the

percentage of correct responding for each 10-trial session during the 60%, 80%, and

100% for three sessions mastery criteria series.

Results. Figure 4 depicts the results for the verbal tacting curricular program for Sandy, Evan, Cyril, and Adam. It is important to note that for this curricular program, a verbal response is required for each individual stimulus card rather than a physical response of selecting a card within a stimulus array. As such, the opportunity for chance responding differs from that of the two previous curricular programs (Experiments 1 & 2).

For Sandy, (Figure 4, top panel) responding during maintenance probes for the 60% correct for 3 sessions mastery criterion immediately decreased to 30% at week 1 and further decreased to zero-level responding across weeks 2 to 4. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to zero-level responding at week 1 and remained at that level across weeks 2 to 4. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series remained at 100% at week 1 and across all subsequent maintenance probe sessions.

For Evan (Figure 4, 2nd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion remained at 70% at week 1. A steady decline was observed across subsequent maintenance probes to a low of 30% correct. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to 30% at week 1. A similar decreasing trend was observed and responding returned to zero-levels by week 3. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series slightly decreased to 90% at week 1. A similar decreasing trend was observed; however, responding had only decreased to 70% by week 4.

For Cyril (Figure 4, 3rd panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion immediately decreased to 30% at week 1 and ranged from 30% to 70% across weeks 1 to 4. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series immediately decreased to 50% at week 1. However, responding returned to 100% (a level observed during the last acquisition session) at week 2 and remained at this level for subsequent maintenance probes. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series decreased to 80% at week 1.   Responding during weeks 1 to 4 ranged from 70% to 100% correct.

For Adam (Figure 4, bottom panel), responding during maintenance probes for the 60% correct for 3 sessions mastery criterion immediately decreased to 50% at week 1 and ranged from 40% to 70% across weeks 1 to 4. Responding during maintenance probes for the 80% correct for 3 sessions mastery criterion series remained at 80% at week 1. Responding then decreased to 60% for all subsequent maintenance probes. Finally, responding during maintenance probes for the 100% correct for 3 sessions mastery criterion series remained at 100% at week 1 and across all subsequent maintenance probe sessions.

Experiment 3 Discussion. Across all participants, 12 maintenance probe data series are amenable to visual inspection. This includes four maintenance probe data series for each of the three maintenance criteria levels (i.e., 60%, 80%, and 100% correct responding for three sessions) (see Appendix J).

For the 60% correct responding for three sessions criterion, two of the four maintenance probe data series appeared to be maintaining at or slightly below criterion

levels at week 1. Across weeks 2 and 4, patterns of responding appeared idiosyncratic. That is, increasing trends comparable to acquisition patterns were observed (one of four participants), steady responding comparable to acquisition levels were observed (one of four participants), drastically decreasing trends toward zero-levels were observed (one of four participants), and zero-level responding was observed (one of four participants).

For the 80% correct responding for three sessions criterion, one of the four maintenance probe data series appeared to be maintaining at criterion levels at week 1. For two of the four data series, responding had decreased to zero or near zero levels at week 1 and all subsequent maintenance probes. The remaining data series shows a decrease to   The remaining two data sets shows a decrease to 50% at week 1 but performance at 100% for subsequent maintenance probes. It is unclear what contributed to the initial decrease in responding. However, it is interesting to note that for this participant, 100% correct responding had previously been demonstrated during the final acquisition session. This may have contributed to higher levels of responding observed across subsequent maintenance probes.

For the 100% correct responding for three sessions criterion, three of the four maintenance probe data series maintained at 100% at week 1. The results are findings are identical to results from Experiments 1 and 2. Across all four maintenance probe data series, responding remained above 70% across all 4 weeks. For two participants, responding below 100% was never observed during any maintenance probe sessions. Overall, data indicate that the 80% correct for three sessions in a row mastery criterion resulted in early maintenance at criterion levels for one of the four participants, significant deterioration for two of the four participants, and complete deterioration to

chance-levels for the remaining participant by week 1. These findings are similar to the results of Experiments 1 and 2. Later maintenance probe data show that for two participants, responding remained at or above 60%. For the remaining two participants, responding was observed at 20% or below. These data suggest that this criterion, alone, may result in maintained or moderately decreased levels of responding for some individuals and result in little to no maintenance for an approximately equal numbers of individuals. These findings are similar to that of Experiment 2 and are only slightly better than the findings of Experiment 1. Similar to Experiments 1 and 2, across all four comparisons, responding in the 80% for three sessions criterion series was observed to remain consistently higher than the 60% for three sessions criterion series in two cases. In addition, these results further suggest that utilizing a 100% correct responding for 3 sessions criterion may be more effective for promoting maintenance of responding.

Experiments 1, 2, & 3 Discussion. Across Experiments 1, 2, and 3, 36 maintenance probe data series are amenable to visual inspection. This includes 12 maintenance probe data series for the 60% correct responding for three sessions criterion, 12 maintenance probe data series for the 80% correct responding for three sessions criterion, and 12 maintenance probe data series for the 100% correct responding for three sessions criterion.

Overall, the 80% correct responding for three sessions criterion was observed to be correlated with patterns of responding consistently above those observed during the 60% for three sessions criterion for only three of twelve comparisons. Further, the 80% correct responding for three sessions criterion resulted in maintained responding at or slightly below criterion-levels at week 1 for only 5 of 12 data series.

In comparison, the 100% correct responding for three sessions criterion was observed to be correlated with patterns of responding consistently above those observed during the both the 60% and 80% criteria for 11 of 12 comparisons. Further, the 100% correct responding for three sessions criterion resulted in maintained responding at week 1 at 100% for 5 of 12 data series, 90% for 2 of 12 data series, and 80% for the remaining 5 of 12 data series. That is, performance below 80% accuracy was never observed across all experiments at one week following acquisition at this mastery criterion. For the 80% correct responding for three sessions criterion, performance at or below 50% at week 1 was observed 67% (8 of 12 observations) of the time.

However, several limitations that hinder the formulation of strong conclusions regarding the 80% correct for three consecutive sessions mastery criterion. First, the vast number of different individuals, settings, curricular programs, and unique procedures with which this criterion is utilized is substantial and warrants a large number of direct and slightly-altered empirical replications and extensions. As such, Experiment 4 provided an additional evaluation of this mastery criterion and its impact on maintenance.

Second, a potential limitation of Experiments 1, 2, and 3 was that following the demonstration of each mastery criterion with a given set of stimuli, training with that set of stimuli was entirely halted until maintenance probes were conducted. For most participants, the 100% correct for three sessions criterion often resulted in the highest number of training sessions required to meet mastery. Thus, during the final sessions within the acquisition phase, only one set of stimuli (two or three stimuli total) was being taught in comparison to three sets of stimuli (six or nine stimuli total). Although other acquisition skills were presumably being taught when participants returned to the

classroom, it is possible that massed-trial-training of fewer acquisition stimuli within the experimental context may have contributed to the higher levels of maintenance observed for the 100% correct for three sessions criterion series. As such, Experiment 4 introduced training on additional stimulus sets after a given target training set met mastery criterion levels of accuracy. That is, participants were always being exposed to three sets of stimuli (nine stimuli total) at all times during the acquisition phase.

Third, during the survey portion of this study, many of the respondents indicated that the use of the 80% correct responding for three sessions in a row criterion was utilized in combination with other variables to plan for maintenance including training across two or more therapists, training across two or more environments, interspersing targets with previously mastered targets, and thinning the schedule of reinforcement. Another potential limitation of these experiments is that the effectiveness of utilizing these additional strategies in combination with the 80% for three sessions in a row criterion was not empirically tested; however, some of these variables were inadvertently incorporated to a small extent. Although the author conducted approximately 80% of sessions, 20% of training sessions were conducted across two other therapists. In addition, on four occasions, a training session was conducted within an empty corridor in the school. This was done to accommodate the rare occasion during which two participants were only able to be pulled from class for session at the same time. As such, these procedural variations were controlled for in Experiment 4. That is, all sessions were conducted by only one therapist and only within the designated experimental room setting.

Fourth, a potential limitation of these experiments may be that during maintenance probe sessions, contingent reinforcement was not provided. Rather, access to preferred stimuli was provided non-contingently to reduce the probability of non-compliant behavior associated with escape from demands. It is possible that lower levels of accurate responding during later maintenance probe sessions may have been a result of a "non-contingent schedule of reinforcement" or prolonged exposure to extinction. That is, the most salient indicator of skill maintenance should be the first maintenance probe session. In an ideal clinical setting, once a certain skill is mastered, the target is quickly moved into maintenance training. This would involve regularly-scheduled probes and supplemental teaching with access to contingent reinforcement. Thus, Experiment 4 only conducted a maintenance probe at one week following acquisition at a given mastery criterion level.

Finally, during the survey portion of this study, most respondents indicated utilizing an 80% mastery criterion across multiple sessions (73%) and across multiple sessions with additional variables (55%). Thus, Experiments 1, 2, and 3 evaluated this percentage level of performance. In addition, this level was compared to 100% accuracy and 60% accuracy. Although these levels were arbitrarily selected as ceiling and floor levels for experimental purposes, very few individuals reported utilizing these percentage levels in clinical settings. Only 9% of individuals reported utilizing a 100% criterion across multiple sessions and 8% reported utilizing a 100% criterion across multiple sessions with additional variables. However, 39% of respondents indicated utilizing a 90% criterion across multiple sessions and 26% utilized the same percentage across multiple-sessions with additional variables. As such, Experiment 4 provided an

evaluation of the often-most reported criterion of 80% correct as well as the second-most reported criterion of 90%. It is possible that while 80% correct across three sessions did not reliably produce maintained responding, 90% correct across three sessions may produce similar acceptable levels of maintained responding as those observed for the 100% criterion. Thus, Experiment 4 empirically evaluated the extent to which an 80%, 90%, and 100% correct responding across three sessions criterion resulted in the maintenance of skills.

Experiment 4

Experiment 4 served as a replication and extension of Experiment 3. It included the same four individuals who passed ABLA-R Level 6. Experiment 4 only differed from Experiment 3 in that after a given set of stimuli were mastered (per the designated criterion level), a new set of dummy stimuli were introduced and taught. In addition, all sessions were conducted by one trainer in one experimental setting (controlling for the influence of other variables). Maintenance probes were only conducted at one week following performance at the respective mastery criterion levels. Finally, Experiment 4 compared the maintenance of skills across 80%, 90%, and 100% correct responding for three consecutive sessions.

Method.

Participants and setting. Four individuals who passed ABLA-R level 6 (Evan, Cyril, Sandy, and Adam) participated in Experiment 4. The experimental setting was identical to Experiments 1, 2, and 3.

Design and procedures. The experimental design and procedures utilized in Experiment 4 were identical to those described in Experiment 1. However, this

experiment involved an evaluation of the maintenance of skills following teaching until a mastery criterion of 80% correct responding for three consecutive sessions was met as compared to both a mastery criterion of 100% correct responding for three consecutive sessions and a mastery criterion of 90% correct responding for three consecutive sessions.

Each of the four participants who passed level 6 on the ABLA-R were trained with three different sets of discriminative stimuli, one set randomly assigned to each of the three conditions (i.e., 100%, 90%, and 80% mastery criteria levels) for one category of skills. After performance at the designated mastery criteria level was observed for a given set, training on the set was discontinued and training on a novel dummy set of stimuli began. This continued until mastery performance levels were met across all three sets of target stimuli (i.e., 80%, 90%, and 100% correct for three sessions). For each participant, maintenance probe data were collected 1 week immediately following performance at the assigned mastery criteria level for a given target stimulus set.

Inter-observer agreement and procedural integrity. IOA and procedural integrity data were collected as described in Experiment 1. IOA data was collected for 53% of baseline sessions, 38% of acquisition sessions, and 50% of maintenance probe sessions. This included dummy set sessions (not graphed). For all participants, the mean agreement for IOA is 100% across baseline, acquisition, and maintenance probe sessions. Procedural integrity data was collected for 38% of acquisition sessions. Across all participants, procedural integrity scores averaged 100%. The mean IOA score for procedural integrity data was 100%.

Curricular program.

Verbal tacting. The curricular program taught in Experiment 4 was identical to the curricular program taught in Experiment 3. However, entirely novel stimuli and target responses were utilized. Three target sets of three pictures were included in this curricular program with each set being randomly assigned to the three mastery criteria levels (i.e., 60%, 80%, or 100%). In addition, two dummy sets of three pictures were included in this experiment. No mastery criteria rules were applied to these two sets and training continued regardless of performance.

Results. Figure 5 depicts the results for the verbal tacting curricular program for Evan, Cyril, Adam, and Sandy. For Evan, (Figure 4, top panel) responding during maintenance probe for the 80% correct for 3 sessions mastery criterion immediately decreased to 0% at one week following the final acquisition session for this stimulus set. Responding during maintenance probe for the 90% correct for 3 sessions mastery criterion series decreased to 40%. Finally, responding during maintenance probe for the 100% correct for 3 sessions mastery criterion series decreased to 70%.

For Cyril (Figure 4, 2nd panel), responding during maintenance probe for the 80% correct for 3 sessions mastery criterion decreased to 20% at one week following the final acquisition session for this stimulus set. Responding during maintenance probe for the 90% correct for 3 sessions mastery criterion series immediately decreased to 0%. Finally, responding during maintenance probe for the 100% correct for 3 sessions mastery criterion series decreased to 70%.

For Adam (Figure 4, 3rd panel), responding during maintenance probe for the 80% correct for 3 sessions mastery criterion decreased to 60% at one week following the final acquisition session for this stimulus set. Responding during maintenance probe for the
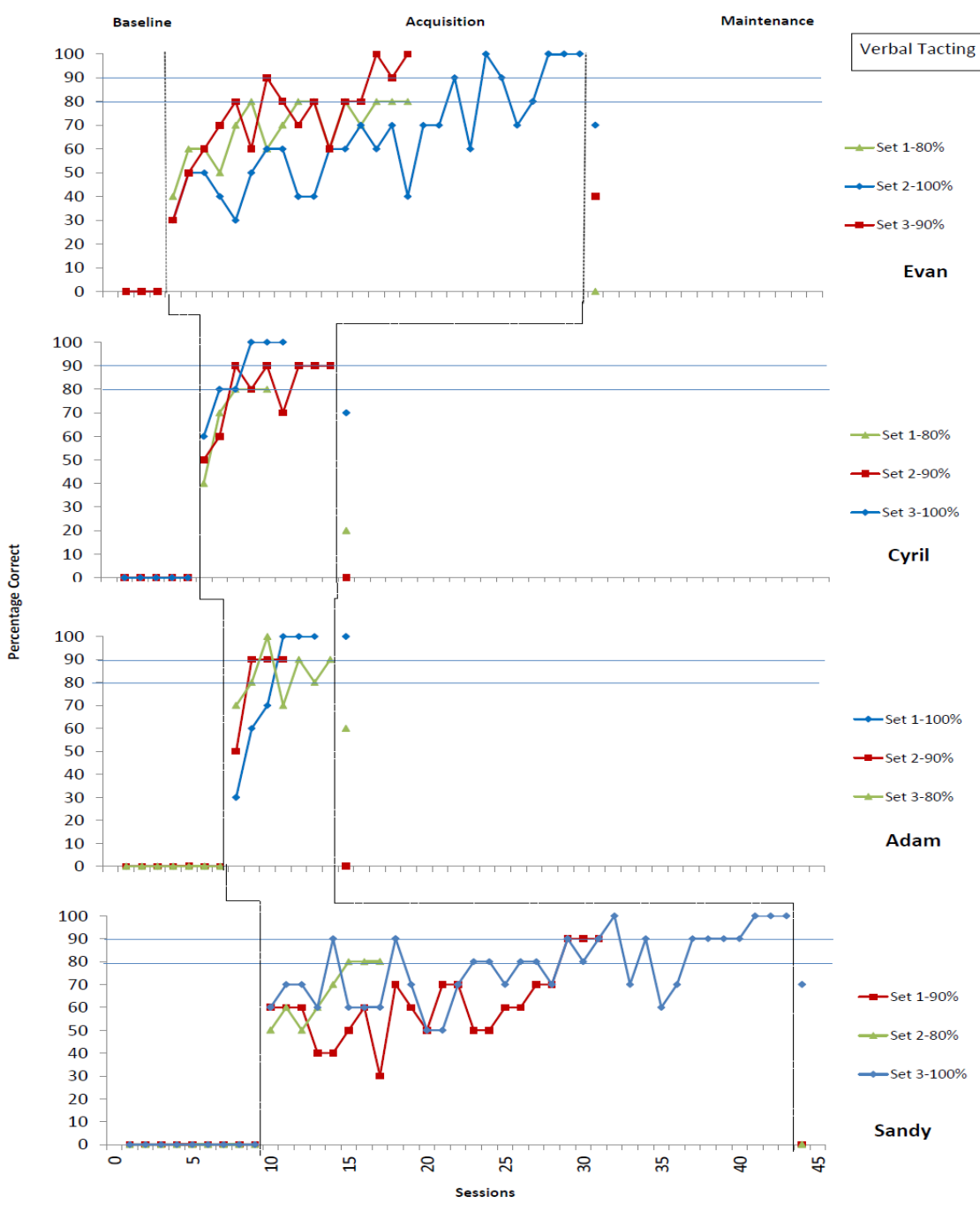
Figure 5. Verbal tacting data across baseline, acquisition and maintenance for Evan, Cyril, Sandy, and Adam who passed Level 6 of the ABLA-R. The figure indicates the percentage of correct responding for each 10-trial session during the 60%, 80%, and 100% for three sessions mastery criteria series.

90% correct for 3 sessions mastery criterion series immediately decreased to 0%.

Finally, responding during maintenance probe for the 100% correct for 3 sessions

mastery criterion series remained at 100%.

For Sandy (Figure 4, bottom panel), responding during maintenance probe for the

80% correct for 3 sessions mastery criterion immediately decreased to 0% at one week

following the final acquisition session for this stimulus set. Responding during

maintenance probe for the 90% correct for 3 sessions mastery criterion series also

immediately decreased to 0%. Finally, responding during maintenance probe for the

100% correct for 3 sessions mastery criterion series decreased to 70%.

Experiment 4 Discussion. Across all participants, 12 maintenance probe data series are

amenable to visual inspection. This includes four maintenance probe data series for each

of the three maintenance criteria levels (i.e., 80%, 90%, and 100% correct responding for

three sessions) (see Appendix K).

For the 80% correct responding for three sessions criterion, one of the four

participants demonstrated slightly decreased responding at 60% accuracy during the

maintenance probe. One of the four participants demonstrated significantly decreased

responding at 20% accuracy. For the remaining two participants, responding had entirely

deteriorated to zero-levels after one week post-acquisition sessions.

For the 90% correct responding for three sessions criterion, one of the four

participants demonstrated significantly decreased responding at 40% accuracy during the

maintenance probe. For the remaining three participants, responding had entirely

deteriorated to zero-levels after one week post-acquisition sessions.

For the 100% correct responding for three sessions criterion, one of the four participants demonstrated maintained responding at 100% accuracy during the maintenance probe. For the remaining three participants, responding had slightly decreased to 70% accuracy after one week post-acquisition sessions.

In combination with the results of the previous three experiments, these data suggest that the 80% correct for three sessions criterion, alone, may not be sufficient for reliably producing maintained responding following training sessions. In addition, these results further suggest that utilizing a 100% correct responding for three sessions criterion may be more effective for promoting the maintenance of responding. As previously mentioned, performance below 80% accuracy was never observed across Experiments 1, 2, and 3 at one week following acquisition for the 100% criterion. In Experiment 4, responding for three participants was slightly lower (i.e., 70%) at one week post-acquisition. However, the highest performance observed during maintenance probes for the other two mastery criteria was 60% for one participant on the 80% correct stimulus set. Across the remaining 80% and 90% for three sessions series, responding fell at or below 40%. The results of Experiment 4 also tentatively suggest that a 90% correct responding for three sessions criterion may not be sufficient for reliably producing acceptable levels of maintained responding post-acquisition. However, further replications are certainly needed to support this possible conclusion.

Overall, decreases in responding were even more pronounced in Experiment 4 than in Experiments 1, 2, and 3. Although it is empirically unclear why these lower levels were observed, there are several potential explanations. First, Experiment 4 was conducted by only one experimenter in one experimental setting. It is possible that these

additional variables utilized by many practitioners contribute to higher level of responding. Future research should systematically evaluate the extent to which the planned addition of these procedures results in higher levels of maintained responding. Still, it is worth noting that with the 100% correct across three sessions criterion, decreases in response accuracy were not as pronounced. Future research should evaluate the extent to which the planned addition of these procedures results in optimal levels of maintained responding across longer durations of time post-acquisition.

Second, unlike the previous experiments, Experiment 4 included additional dummy training sets of stimuli. It is possible that the increased number of total stimuli taught impacted the maintenance of responding related to those stimulus sets. Future research may evaluate the ideal number of acquisition programs and stimuli to be taught at a given time and the factors that contribute to selecting that number of targets.

Third, the type of stimuli utilized in Experiment 4 were visual stimuli that are more likely to be taught in academic settings. That is, stimuli included common mathematical symbols, states, and geographical features. As such, it is possible that some features across the stimuli required more complex and finite discriminations. For example, a multiplication sign and a plus sign differ only with respect to orientation and not with respect to the color or shape. Additional research related to refining the levels of the ABLA-R to include additional variations and combinations of different types of visual discriminations might be useful for guiding the selection of appropriate stimuli to teach.

General Discussion

The current study was designed to evaluate the extent to which the most commonly-reported mastery criterion rules resulted in skill maintenance in the context of behavioral skills training with children diagnosed with autism and intellectual disabilities. This mastery criterion was compared to three other mastery criteria (i.e., 100% correct responding for three sessions, 60% correct responding for three sessions, and 90% correct responding for three sessions). Three curricular programs were involved in this study, with comparisons made across eight children. A multiple-baseline design combined with an alternating treatments design was utilized to compare the four mastery criteria for each of the curricular programs assigned to the participants. These curricular programs were assigned to each of the participants based upon the results of the ABLA-R phase of the current study. In general, the results suggest that an 80% correct responding for three sessions mastery criterion utilized by a majority of practitioners (as determined during the survey of clinical practices phase of the current study) may not be sufficient for maintenance of performance for some individuals.

Findings suggest that the additional acquisition sessions required to meet a 100% correct responding for three sessions criterion produced higher levels of maintenance than the 60% for three sessions criterion, the 80% for three sessions criterion, and the 90% for three sessions criterion. For 12 of the 16 data sets, the 100% correct responding for three sessions mastery criterion resulted in the greatest number of acquisition sessions and discrete-trial presentations (Appendix L). The 100% accuracy for three sessions criterion not only provided a greater number of sessions and discrete trial presentations, but also provided a greater number of reinforcer presentations during acquisition.

Lerman and Iwata (1996) conducted a review of extinction literature and concluded that basic human and nonhuman investigations have demonstrated that the number of reinforcer presentations during skill acquisition, the number of discrete trial presentations, and the length of skill acquisition training may affect resistance to extinction. In addition, the authors suggested that these variables may be manipulated to increase the maintenance of responding. Several sources have indicated that as the number of reinforcer presentations and length of skill acquisition increases, an increase in resistance to extinction is observed (e.g., Cooper, Heron, & Heward, 1987; Grant & Evans, 1994; Kazdin, 1994; Mercer & Snell, 1977). In addition, several sources have suggested that the use of partial-reinforcement schedules (e.g., Lewis, 1960; Mackintosh, 1974) and the use of partial-reinforcement schedules combined with long skill acquisition training (Uhl & Young, 1967) can produce an increased resistance to extinction.

Lerman and Iwata (1996) stated that several studies have demonstrated that the use of a partial-reinforcement schedule following a period of exposure to a continuous reinforcement schedule results in an increase in resistance to extinction (Nation & Boyajian, 1980; Pittenger, Pavlik, Flora, & Kontos, 1988). The current study involved a continuous reinforcement schedule in the form of praise and a variable-ratio schedule in the form of access to preferred stimuli throughout the entire acquisition phase. However, the purpose of this study was not to evaluate the effectiveness of utilizing a systematically thinned-scheduled of reinforcement following skill acquisition on the maintenance of responding. Rather, the focus of this study was to evaluate the use of specific mastery criteria during skill acquisition in the absence of planned schedule-thinning following acquisition.

Weatherly (2008) conducted an experiment that addressed this issue and concluded that overlearning using a thinned partial-reinforcement schedule following a continuous schedule of reinforcement during acquisition produced performance at a 90% mastery criterion more effectively than both overlearning (continuous reinforcement) and a control condition devoid of overlearning following skill acquisition. In that study, although the continuous reinforcement schedule resulted in a greater number of reinforcer presentations than the thinned-schedule of reinforcement schedule, it did not result in greater skill maintenance.

In the current study, the 100% for three sessions mastery criterion produced a greater number of reinforcer presentations than the 90%, 80% and 60% mastery criteria and also produced the highest levels of performance during maintenance probes. This finding is consistent with results of previous research. However, although the 80% for three sessions mastery criterion produced a greater number of reinforcer presentations than the 60% criterion (as well as the 90% criterion relative to the 80% criterion) it did not always result in superior performance during maintenance probes. In addition, scatter plot data suggest that the sessions required to meet mastery did not correlate with maintenance performance (Appendix M).

One potential explanation for this result is that maintenance probe sessions involved a non-contingent delivery of preferred stimuli to guard against non-compliance and non-responding. It is possible that the delivery of preferred stimuli could potentially follow correct responses and lead to a greater chance for response maintenance. It is also possible that the delivery of preferred stimuli could potentially follow incorrect responses and lead to a decrease in correct responding. Although this procedure was identical

across all four mastery criteria series, it is possible that a greater number of correct responses were inadvertently reinforced during the 100% mastery criterion maintenance probes than the 90%, 80%, and 60% mastery criterion maintenance probes.

Given the large number of discrete trial presentations required for the current investigation relative to the number of discrete trial presentations observed during typical instruction within the classroom combined with the number of non-compliant behaviors observed within the typical environment, it was determined that a non-contingent reinforcement schedule would be necessary during baseline and maintenance probe sessions to decrease the probability of non-responding and non-compliant behaviors. However, it is worth noting that during the maintenance probe sessions, sessions typically took between less than 1 minute and 2 minutes to complete. As such, the probability of inadvertent reinforcement following either correct or incorrect responses during this phase of the study was relatively low across all three mastery criteria series. For the majority of sessions, the delivery of preferred stimuli followed the completion of an entire 10-trial block.

Weatherly (2008) provided an inquiry into what component of the response was maintained and what component can be said to have undergone extinction. The current study utilized the same requirement of some type of operant response for every trial (i.e., trials with non-responding were not included in the data presentation). Weatherly suggested that, as such, behavior did not extinguish during the maintenance probes in that the participants always made an operant response. The author noted that the variability of the correctness of the response increased during maintenance testing. It was stated that incorrect responses observed included one of the responses that had previously been

reinforced in the presence of a certain discriminative stimulus (e.g., "Do this") or

responses within the same class. Similar patterns of errored responding were observed

during maintenance probe sessions in the current study across curricular programs.

Unlike the Weatherly (2008) study, the current study also involved a curricular

program that required a verbal response (i.e., the verbal tacting program). As in the

Weatherly study, the types of errors observed during this curricular program involved

verbal responses that were previously reinforced in the presence of the particular

discriminative stimulus (i.e., "What is this?"). Thus, these kinds of errors were not novel

responses from other unrelated response classes which demonstrates the maintenance of

response differentiation (Weatherly). The author noted that, as suggested by Lerman and

Iwata (1996), when a behavior is exposed to extinction conditions other behaviors from

that response class will emerge. In the current study, this pattern of responding was also

observed in that participants also errored by responding "I don't know", "help me", or "I

can't remember". It is presumed that these verbal responses had previously been

reinforced in the presence of similar or identical discriminative stimuli and, as such, is

consistent with Weatherly's conjectures. Interestingly, participants also errored by

responding with a portion of the discriminative stimulus presented during the other

curricular programs. For example, during the receptive identification program, the

participant was told "point to lemur". During maintenance probe sessions within the

verbal tacting curricular program, the participant was asked "What is this?" and would

respond "lemur" even though the picture was of a different animal.

Another notable difference between the current study and the Weatherly (2008)

relates to the levels of correct responding observed during Weatherly's control condition.

Prior to the implementation of the overlearning (continuous reinforcement), overlearning

plus thinned schedule of reinforcement, and control conditions, the author reported

teaching skills to 90% accuracy for two sessions in a row. Although the higher levels of

maintained responding during the two overlearning conditions may be attributed to the

implementation of these additional variables following reaching mastery, responding in

the control condition always remained above 70%, even during maintenance probes

conducted between 40 and 80 days following demonstrating mastery. In the current

study, responding remained at or above 70% during preliminary maintenance probes for

the 100% for three sessions mastery criterion sets. However, immediate decreases in

responding below 70% were observed for the 90%, 80% and 60% mastery criteria sets,

even after just one week following skill acquisition.

One potential explanation for this difference in skill maintenance across the two

studies may relate to the kinds of participants involved in each of the investigations. The

Weatherly (2008) study was conducted with participants involved in a public preschool

autism program. In the description of the settings, the author noted that the staff:child

ratio within the classroom was 1:1 and the tutors running the program were

undergraduate and graduate students from a local university. In addition, a certified

special education teacher, two classroom aides with bachelor's degrees in psychology and

training in behavior analysis, four graduate students in behavior analysis, and a behavior

analysis faculty member were involved. The individuals within this classroom attended

school for three hours a day, five days a week. The author noted that this preschool

classroom was one of three preschool classrooms that utilized the Autistic Impaired

Preschool Program within this school. This program involved the regular use of discrete-trail training on an individual level within individual study carols.

The classroom of the participants involved in the current study differed significantly. That is, the staff:child ratio within the classroom is approximately 1:5 and the staff and aides are of various educational backgrounds. The classroom includes two teachers, one of which is currently obtaining a master's degree in behavior analysis. The remaining aides within the classroom did not have training in behavior analysis. In addition, although the use of discrete trial training and a token-economy was common within classroom, students were taught in a group format. When individual training was conducted with each student, other students were present and attended to at the same table by a single aide or teacher. As such, the participants in the Weatherly (2008) study may have had a significantly longer history of responding in conditions similar to that utilized within the research. The tutors within the classroom were involved in conducting sessions and likely had a substantial history of providing reinforcement for accurate responding in a discrete trial format. While the study was being conducted, the participants were presumably being reinforced for accurate responding across a variety of programs in the presence of the same tutors. As such, maintenance of responding in during the control condition of the study may be related to the stimulus control of the conditions and tutors as well as a continued delivery of reinforcement for accuracy of responding, in general, in the typical operations of the classroom. In the current study, sessions were conducted by this author and two research assistants not involved in the operation of the typical classroom.

The lower levels of maintenance observed during the current study may be attributed to a shorter period of reinforcement history under similar stimulus conditions. It could be argued that the conditions described in the Weatherly (2008) study are more comparable to the conditions under which behavior analytic services are provided by individuals in the practice domain of behavioral psychology than those of the current study. This point notwithstanding, individuals receiving behavior analytic services may differ with respect to the number of hours of services received per week. Some individuals may receive 15 hours of behavior analytic services per week, whereas others may receive significantly less. The Weatherly study did not report if the participants involved also received behavior analytic services outside of the pre-school program. Presumably, many individuals receiving behavior analytics services may attend schools in which a formal behavior analytic curriculum provided by individuals trained in behavior analysis is not commonplace. Further, anecdotal observations on the part of the current investigator and discussions with professionals from the education field suggest that the use of the 80% criterion for three sessions in a row is not specific to the field of behavior analysis. As such, the current research is still warranted. Future research should be conducted evaluating the extent to which the results of this study compare to results obtained within the context of a typical behavior analytic service delivery system (as determined by a survey of clinical practices).

Another potential explanation for the lower levels of maintained responding observed during the 60% and 80% for three sessions criteria series in the current study as compared to the control condition of the Weatherly (2008) study may relate to the specific mastery criterion utilized. Weatherly reported teaching skills to 90% accuracy

across two sessions criterion. This criterion alone may be more effective at producing maintained responding than an 80% for three sessions criterion. Interestingly, Weatherly reported that the setting in which the research was conducted utilized either an 80% accuracy for three consecutive sessions mastery criterion or a 90% accuracy for three consecutive sessions mastery criterion. However, the final experiment of the current study provided preliminary data that suggest the inclusion of the 90% for three consecutive sessions mastery criterion, alone, did not produce maintained responding at the levels observed by Weatherly. Future research should evaluate the extent to which alternate variations with respect to the mastery criterion such as other percentages, numbers of sessions, and additional variables (e.g., multiple therapists, environments, time between sessions, schedule-thinning, and task interspersal) impact maintained responding.

As previously mentioned, during the survey portion of this study, many of the respondents indicated that the use of the 80% correct responding for three sessions in a row criterion was utilized in combination with other variables to plan for maintenance including training across two or more therapists, training across two or more environments, interspersing targets with previously mastered targets, and thinning the schedule of reinforcement. One potential limitation of this study is that effectiveness of utilizing these additional strategies in combination with the 80% for three sessions in a row criterion was not empirically tested.

However, it is important to note that during the first three experiments of this study, some of these variables were incorporated to a certain extent. That is, even though the author conducted approximately 80% of sessions, 20% of training sessions were

conducted across two other therapists. During Experiment 4, these additional variables were not present and responding was observed to deteriorate to a greater extent. In addition, at some points, mastery criterion sessions were being conducted with some targets while some other targets remained in the acquisition phase. Although mastered targets were not interspersed with acquisition targets on a trial-by-trial basis, trial blocks of mastered items were sometimes conducted back-to-back with acquisition trial blocks. Finally, during acquisition, reinforcement in the form of praise was provided on a continuous schedule; however, reinforcement in the form of access to tangible items was provided on an intermittent schedule of reinforcement (i.e., a variable-ratio 3 schedule of reinforcement). Although this does not represent a systematic thinning of the schedule of reinforcement, the variable nature of the intermittent reinforcement schedule should promote maintenance of responding during extinction conditions or, in this case, non-contingent reinforcement conditions.

Another potential limitation of the current study is that it involved only 8 participants (16 data sets) and 3 different types of skills. When considering the extent to which mastery criterion rules such as 80% or above for three sessions in a row are utilized within the field of applied behavior analysis, additional research is necessary to evaluate the utility of these kinds of rules across a larger sample of individuals and skills. During the survey portion of the current study, a small number of participants commented that the mastery criterion they used sometimes differed depending on the type of skills being taught and the individual learner. Different types of skills may require different levels of response accuracy. For example, a skill such as learning to look both ways before crossing the street should be taught to 100% accuracy to avoid danger to the

individual. Other skills may not require responding at 100% accuracy. However, the current study represents a preliminary investigation into a large-scope issue within the applied and service delivery domains of behavior analysis. As is imperative in the field of behavior analysis, the strategies utilized should be considered on an individual basis and should be derived from evidence-based literature. The participants involved in the current study represent a small sample of individual learners with varying levels of functioning as determined by the ABLA and the skills taught are common skills taught within the field. The current study was intended to be a starting point for further investigations in a much-needed line of research.

Another potential limitation of the current study is the possibility for individuals to respond at higher levels of accuracy during the skill acquisition phase than the percentage levels randomly assignment for that skill set. For example, responding could have occurred at 100% or above for three consecutive sessions even though a particular skill set was randomly assigned to the 60% for three sessions criterion. However, this did not occur for any of the current data sets. This result is certainly fortunate from an empirical standpoint. Even if this were to have occurred within the current investigation, it would be representative of the potential for these patterns to occur in practice. Ultimately, the current investigation does not aim to determine the extent to which behavior occurring at exactly 80% accuracy across three sessions maintains. Rather, the current investigation aims to determine the extent to which an arbitrary rule of 80% accuracy or above results in the maintenance of skills, regardless of the exact patterns of responding obtained by implementing this rule. Even though this mastery criterion rule might result in skill maintenance for some individuals, overall, a mastery criterion rule

that results in skill maintenance for most individuals is preferable. This is especially true

if that rule is to be adopted as a standard within the field of behavior analysis and written

into the respective institutional policies and insurance requirement standards.

Another potential limitation of the current study is that it only evaluated mastery

criteria based on a percentage of correct responses across sessions. Additional research is

needed to evaluate differing mastery criteria based on number of trials correct and a rate

of responding per unit of time. Studies of this nature could easily be applied to different

performance evaluation systems such as that of precision teaching (e.g., Lindsley, 1992)

which utilizes fluency aims rather than mastery criteria.

Overall, the use of mastery criteria as a measure of skill acquisition is pervasive

across a variety of domains including behavior analysis and education. The use of

evidence-based strategies is integral within the professional service delivery domain of

applied behavior analysis. The wide-spread adoption of particular teaching procedures

such as the use of an 80% accuracy across three session mastery criterion may represent a

drift away from evidence-based practice. Although the current study may provide some

preliminary evidence that this mastery criterion rule alone is ineffective for promoting the

skill maintenance for some individuals, a large amount of additional research is needed to

replicate and extend upon these findings. Although the current study was not designed to

statistically assess the contribution of multiple variables on observed maintenance

performance, scatterplot data show a great deal of variability and lack of clear order

(Appendix M). The assigned mastery criterion level, sessions required to acquire mastery

criterion, performance on the last acquisition session, and average performance on the

last three acquisition session do not appear to trend significantly with maintenance performance.

Findings from this study suggest that the additional variables utilized by many practitioners such as applying the mastery criteria across multiple trainers, days, and settings may contribute to higher levels of maintained responding. As such, future research should evaluate the extent to which the 80% mastery criterion is sufficient for producing acceptable levels of maintenance performance when combined with these additional variables. Additional research should also compare the relative benefits of utilizing various combinations of percentage levels, number of sessions, days, trainers, settings and other variables in order to determine the most efficient and effective strategies for producing clinically significant and sustainable increases in skill performance beyond our clinical involvement. ABA is currently viewed as evidence-based and is the gold standard treatment for individuals with Autism and evaluations of standard practices such as those suggested are imperative to maintaining that status.

References

Anderson, S., Taras, M., and O'Malley Cannon, B. (1996). Teaching new skills to children with autism. In *Behavioral intervention for young children with autism: A manual for parents and professionals* (eds. C. Maurice, G. Green, and S.C. Luce), pp. 181–194. Austin, Tex.: PRO-ED.

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

Carr, J. E., Nicolson, A. C, & Higbee, T. S. (2000). Evaluation of a brief multiple stimulus assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33,* 353-357.

Cummings, A. R., & Carr, J. E., (2009). Evaluating progress in behavioral programs for children with autism spectrum disorders via continuous and discontinuous measurement. *Journal of Applied Behavior Analysis, 42,* 57-71.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis.* Columbus, OH: Merrill.

DeWiele, L., Martin, G., Martin, T. L., Yu, C. T., & Thomson, K. (2011). The Kerr Meyerson Assessment of Basic Learning Abilities Revised : A self-instructional manual (2nd ed.). St. Amant Research Centre, Winnipeg, MB, Canada. Freely available from http://stamantresearch.ca/abla/.

Dunlap, G. (1984). The influence of task variation and maintenance on the learning and affect of autistic children. Journal of Experimental Child Psychology, 37, 41–64.

Freeland, J. T., & Noell, G. H. (2002). Programming for maintenance: An investigation of delayed intermittent reinforcement and common stimuli to create

indiscriminable contingencies. *Journal of Behavioral Education, 11,* 5-18.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis.* New York: Harper Collins.

Johnston, J. M. & O'Neill, G. (1973). The analysis of performance criteria defining course grades as a determinant of college student academic performance. *Journal of Applied Behavior Analysis*, 6, 261-268.

Kazdin, A. E. (1994). *Behavior modification in applied settings* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Kerr, N., Meyerson, L., & Flora, J. A. (1977). The measurement of motor, visual, and auditory discrimination skills. *Rehabilitation Psychology*, *24*(3), 95–112.

Leaf, R., and McEachin, J., eds. (1999) *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism.* New York: DRL Books.

Lerman, D. C., & Iwata, B. A. (1996). Developing a technology for the use of operant extinction in clinical settings: An examination of basic and applied research. *Journal of Applied Bheavior Analysis, 29,* 345-382.

Lewis, D. J. (1960). Partial reinforcement: A selective review of the literature since 1950. *Psychological Bulletin, 57,* 1-28.

Lindley, O. R. (1992). Precision Teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25,* 51-57.

Love, J. R., Carr, J. E., Almason, S. M., & Ingebor Petursdottir, A. (2009). Early and intensive behavioral intervention for autism: A survey of clinical practices. *Research in Autism Spectrum Disorders, 3,* 421-428.

Luiselli, J. K., Russo, D. C., Christian, W. P., & Wilczynski, S. P. (2008). Skill acquisition, direct instruction, and educational curricula. In J. K. Luiselli (Ed.), Effective practices for children with autism (p. 196). New York, NY: Oxford University Press.

Mackintosh, N. J. (1974). *The psychology of animal learning.* New York: Academic Press.

Martella, R. C., Nelson, R. & Marchand-Martella, N. E. (1999). *Research methods: Learning to become a critical research consumer.* Needham Heights, MA: Allyn & Bacon.

Martin, G. L., Thorsteinsson, J. R., Yu, C. T., Martin, T. L., & Vause, T. (2008). The assessment of basic learning abilities test for predicting learning of persons with intellectual disabilities: a review. *Behavior Mo. dification*, *32*(2), 228–47. doi:10.1177/0145445507309022

McEachin, J. J., Smith, T., & Lovaas, 0. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation, 97,* 359-372.

Mercer, C. D., & Snell, M. E. (1977). *Learning theory research in mental retardation: Implications for teaching.* Columbus, OH: Merrill.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, *7*(4), 647–653.

Najdowski, A. C., Chlingaryan, V., Bergstrom, R., Granpeesheh, D., Balasanyan, S., Aguilar, B., & Tarbox, J. (2009). Comparison of data-collection methods in a behavioral intervention program for children with pervasive developmental

disorders: A replication. *Jounral of Applied Behavior Analysis, 42,* 827-832.

Nation, J. R., & Boyajian, L. G. (1980). Continuous before partial reinforcement: Effect on persistence training and resistance to extinction in humans. *American Journal of Psychology, 93,* 697-710.

Pereira, J. A., & Winton, S. W. (1991). Teaching and remediation of mathematics: A review of behavioral research. *Journal of Behavioral Education, 1,* 5-36.

Pittenger, D. J., Pavlik, W. B., Flora, S. R., & Kontos, J. (1988). Analysis of the partial reinforcement extinction effect in humans as a function of sequence of reinforcement schedules. *American Journal of Psychology, 101,* 371-382.

Rapp, J. T., & Gunby, K. (2016). Task interspersal for individuals with autism and other neurodevelopmental disorders. *Jounral of Applied Behavior Analysis, 49,* 730-734.

Smith, T. (1999). Outcome of early intervention for children with autism. *Clinical Psychology: Science and Practice, 6,* 33- 49.

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

Uhl, C. N., & Young, A. G. (1967). Resistance to extinction as a function of incentive, percentage of reinforcement, and number of nonreinforced trials. *Journal of Experimental Psychology, 73,* 556-564.

Weatherly, N. L. (2008). *A comparison of maintenance-training methods for children diagnosed with Autism* (Unpublished doctoral dissertation). Western Michigan University, Kalamazoo, MI.

Williams,W. L. (2015). The Assessment of Basic Learning Abilities and Translational

Research in Behavior Analysis. *International Journal of Behavior Analysis and Autism, 1*, 67-83.

Windsor, J., Piché, L. M., & Locke, P. A. (1994). Preference testing: A comparison of two presentation methods. *Research in Developmental Disabilities*, *15*(6), 439–455.

Appendix A

Email Invitation

Greetings,

You are being invited to participate in a research study about clinical practices related to mastery criteria and maintenance. This study is being conducted by Sarah M. Richling, M.S., BCBA and Larry Williams, Ph.D., BCBA-D, from the Department of Psychology at the University of Nevada, Reno. This survey is being conducted in partial fulfillment of a graduate dissertation project. Data from this survey will be used to develop research aimed at determining effective strategies for integrating mastery criteria into behavior analytic teaching programs.

This survey is to be completed by individuals involved in services in the area of intellectual disabilities and autism. The entirety of the survey must be completed in one sitting and will take approximately 5 minutes to complete. Your participation is greatly appreciated. Please click on the link below to take part in the survey.

https://unrpsych.az1.qualtrics.com/SE/?SID=SV_dgpCHbfJclq8XWZ

Appendix B

Survey Questions

1. What is your certification level?
BCBA
BCBA-D
Not currently certified (Disqualifier)

2. Do you currently provide ABA services, oversee/supervise ABA services, or consult on the delivery of ABA services?
Yes
No – disqualified

3. Is your involvement in ABA services in the autism/intellectual disability area?
Yes
No – disqualified

4. Primary role in Behavior Analysis (select one):
Faculty
Practitioner
Administrator
Student

5. Estimate the percentage of clients in each age group who receive ABA services in which you are involved. (Enter whole numbers between 0 and 100. The total must add up to 100). (From JDT survey)
0-6
7-12
13-17
18-and older
Prefer not to answer

6. Highest degree held (in behavior analysis or related field)?
Masters
Doctorate

7. Year degree earned (in behavior analysis or related field)?

Mastery Criterion Questions
Instructions: For the following questions, please respond based on the procedure you use with the majority of your clients.

8. What is the mastery criterion you use based on?
A certain percentage of correct trials (i.e., session-based) [Go to question set A]
A certain number of correct trials in a row (i.e., trial-based) [Got to question set B]
A rate of responses per unit of time – DISQUALIFIED

QUESTION SET A:
If answer 1 above, the following 2 questions are asked:]
    9. What mastery criterion do you use? (select one)
    A certain percentage of correct trials during one session (i.e., trial block) [Go to question
    set A.1]
    A certain percentage of correct trials across multiple sessions [Go to question set A.2]
    A certain percentage of correct trials across multiple sessions and additional
    variables [Go to question set A.3]

        [QUESTION SET A.1]
        10. What percentage of correct trials per session do you use to determine
        mastery?
        Below 80%
        80%
        81%-89%
        90%
        91%-99%
        100%

        [QUESTION SET A.2]
        11. What percentage of correct trials per session do you use to determine
        mastery?
        Below 80%
        80%
        81%-89%
        90%
        91%-99%
        100%

        12. Across how many sessions in a row do you apply this mastery
        criterion?
        2
        3
        4
        More than 4

        [QUESTION SET A.3]

13. What percentage of correct trials per session do you use to determine mastery?
Below 80%
80%
81%-89%
90%
91%-99%
100%

14. Across how many sessions in a row do you apply this mastery criterion?
2
3
4
More than 4

15. Indicate the additional variables that apply to your mastery criterion (check all that apply)
Across two or more therapists
Across two or more environments
First session (trial block) of the day
Follow-up probe after a certain period of time
Target interspersed with previously mastered targets
Reinforcement schedule thinned
Other (please list):

[QUESTION SET B:]

16. What mastery criterion do you use? (select one)
A certain number of trials in a row during one session (i.e., trial block) [Go to question set B.1]
A certain number of trials in a row across multiple sessions [Go to question set B.2]
A certain number of trials in a row across multiple sessions and additional variables [Go to question set B.3]

[QUESTION SET B.1]
17. What number of trials in a row do you use to determine mastery?
Less than 3
3-5
5-10
More than 10

[QUESTION SET B.2]

18. What number of trials in a row do you use to determine mastery?
Less than 3
3-5
5-10
More than 10

19. Across how many sessions in a row do you apply this mastery criterion?
2
3
4
More than 4

[QUESTION SET B.3]
20. What number of trials in a row do you use to determine mastery?
Less than 3
3-5
5-10
More than 10

21. Across how many sessions in a row do you apply this mastery criterion?
2
3
4
More than 4

22. Indicate the additional variables that apply to your mastery criterion (check all that apply)
Across two or more therapists
Across two or more environments
First session (trial block) of the day
Follow-up probe after a certain period of time
Target interspersed with previously mastered targets
Reinforcement schedule thinned
Other (please list):

23. What is the primary source that contributes to the use of your specific mastery criterion?
Graduate school
Funding source
Regulatory requirements (e.g., IEP)
Supervised experience
Continuing education experience (e.g., workshop)
Employer policies/requirements

Appendix C

Data Sheet

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Date/Observer:	1		
	2		
	3		
Condition:	4		
	5		
	6		
Sessions to Criterion:	7		
	8		
	9		
	10		
	%		

Notes:

Participant #:	**Acquisition Set:**

Appendix D

Procedural Integrity Form

Date:_____/_____/_____Tutor:_____Integrity Data
Checker:_____Child:_____

Program:_____ Set:_____Condition:_____

Child's Performance Score:_____

	Trials									
Tutor's Behavior	1	2	3	4	5	6	7	8	9	10
SD (e.g., attending, SD as written, intonation)										
Child's Response (make sure the tutor correctly documents the child's response)										
Correction (if needed) OR										
Reinforcer/Token (immediate, used preferred reinforcer, correctly followed reinforcement schedule)										

*Tutor's Performance Score:_____IOA:_____

*<u>Procedural-Integrity Score</u>

<u>(Tutor's performance score):</u>

The number of correct behaviors divided by the total number of behaviors (30)

Appendix E

Receptive Identification/Visual Discrimination Curricular Program Sample

Group A: Receptive Identification/Visual Discrimination Procedure Sheet

Student:	Oscar		Teachers:	
			Procedure Checkers:	
Objective:	Child points to a stimulus as directed by the tutor			
Materials:	Colored cards (red, yellow), shapes (heart, triangle), sized circles (big, small)			
Reinforcer:	Use the first 3 reinforcers selected by the child before each session			
Data Collection:	10 *compliant* trials, (+) for correct, (-) for incorrect, DO NOT count noncompliant trials toward 10-trial total, however, still provide prompting			

Stimulus Set	Tutor Presentation: (SD)	Correct Response		Incorrect Response		
		Pupil Behavior	**Tutor Behavior**	**Pupil Behavior**	**Tutor Behavior**	**Mastery Criterion**
1	Tutor sits facing student and establishes eye contact with student. Tutor places the red and yellow cards on table. Tutor says, "Point to red". Use red card, randomly alternating sides, and repeat SD for all 10 trials.	Student points to the red card within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>**100% for 3 consecutive sessions**</u>
2	Tutor sits facing student and establishes eye contact with student. Tutor places the heart and triangle shapes on table. Tutor says, "Point to heart". Use heart shape, randomly alternating sides, and repeat SD for all 10 trials.	Student points to the heart shape within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>**80% or greater for 3 consecutive sessions**</u>
3	Tutor sits facing student and establishes eye contact with student. Tutor places the big and small circles on table. Tutor says, "Point to biggest". Use big circle, randomly alternating sides, and repeat SD for all 10 trials.	Student points to the biggest circle within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>**60% or greater for 3 consecutive sessions**</u>

Appendix F

Receptive Identification/Auditory Discrimination Curricular Program Sample

Group B: Receptive Identification/Auditory Discrimination Procedure Sheet

Student:	Adam		Teachers:	
			Procedure Checkers:	
Objective:	Child points to a stimulus as directed by the tutor			
Materials:	Skill 1: Sets 1-3 pictures (willow, magnolia, succulent, lemur, spookfish, armadillo, gibbon, wallaby)			
Reinforcer:	Use the first 3 reinforcers selected by the child before each session			
Data Collection:	10 *compliant* trials, (+) for correct, (-) for incorrect, DO NOT count noncompliant trials toward 10-trial total, however, still provide prompting			

| Stimulus Set	Tutor Presentation: (SD)	Correct Response		Incorrect Response		
		Pupil Behavior	**Tutor Behavior**	**Pupil Behavior**	**Tutor Behavior**	**Mastery Criterion**
1	Tutor sits facing student and establishes eye contact with student. Tutor places the set 1 cards on table. Tutor says, "Point to (card label)". Randomly rotate between all 3 stimuli	Student points to the correct card within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>100% for 3 consecutive sessions</u>
2	Tutor sits facing student and establishes eye contact with student. Tutor places the set 2 cards on table. Tutor says, "Point to (card label)". Randomly rotate between all 3 stimuli	Student points to the correct card within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>80% or greater for 3 consecutive sessions</u>
3	Tutor sits facing student and establishes eye contact with student. Tutor places the set 3 cards on table. Tutor says, "Point to (card label)". Randomly rotate between all 3 stimuli	Student points to the correct card within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>60% or greater for 3 consecutive sessions</u>

Appendix G

Verbal Tacting Curricular Program Sample

Group B: Verbal Tacting Procedure Sheet

Student:	Cyril		Teachers:	
			Procedure Checkers:	
Objective:	Child points to a stimulus as directed by the tutor			
Materials:	Skill 2: Sets 1-3 pictures (kiwi, ginger, octagon, blowfish, starfruit, rhombus, sloth, dragonfruit, spiral)			
Reinforcer:	Use the first 3 reinforcers selected by the child before each session			
Data Collection:	10 *compliant* trials, (+) for correct, (-) for incorrect, DO NOT count noncompliant trials toward 10-trial total, however, still provide prompting			

| Stimulus Set	Tutor Presentation: (SD)	Correct Response		Incorrect Response		
		Pupil Behavior	**Tutor Behavior**	**Pupil Behavior**	**Tutor Behavior**	**Mastery Criterion**
1	Tutor sits facing student and establishes eye contact with student. Tutor presents 1 card from set 1. Tutor asks, "What is this?". Randomly rotate between all 3 stimuli	Student verbally responds with the correct label for the card presented within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>100% for 3 consecutive sessions</u>
2	Tutor sits facing student and establishes eye contact with student. Tutor presents 1 card from set 2. Tutor asks, "What is this?". Randomly rotate between all 3 stimuli	Student verbally responds with the correct label for the card presented within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>80% or greater for 3 consecutive sessions</u>
3	Tutor sits facing student and establishes eye contact with student. Tutor presents 1 card from set 2. Tutor asks, "What is this?". Randomly rotate between all 3 stimuli	Student verbally responds with the correct label for the card presented within 3 seconds of the SD	Provide immediate SR+ (praise) after every correct response and provide 15-30 seconds access to reinforce on a VR-3 schedule.	Student does not make the response within 3 seconds of the SD	Repeat SD and provide a least-to-most prompting hierarchy until correct response is made	<u>60% or greater for 3 consecutive sessions</u>

Appendix H

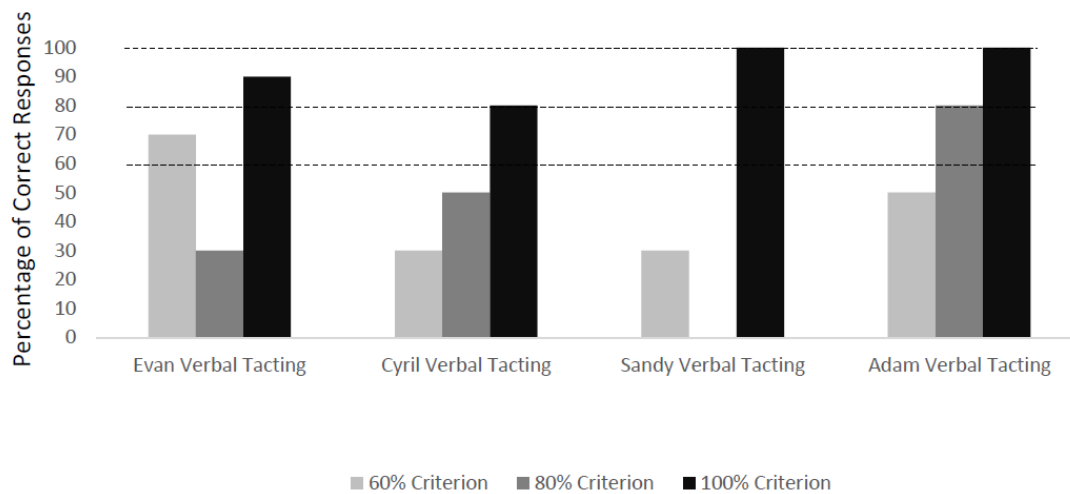Experiment 1: Performance on First Maintenance Probe Session (10-trial block)
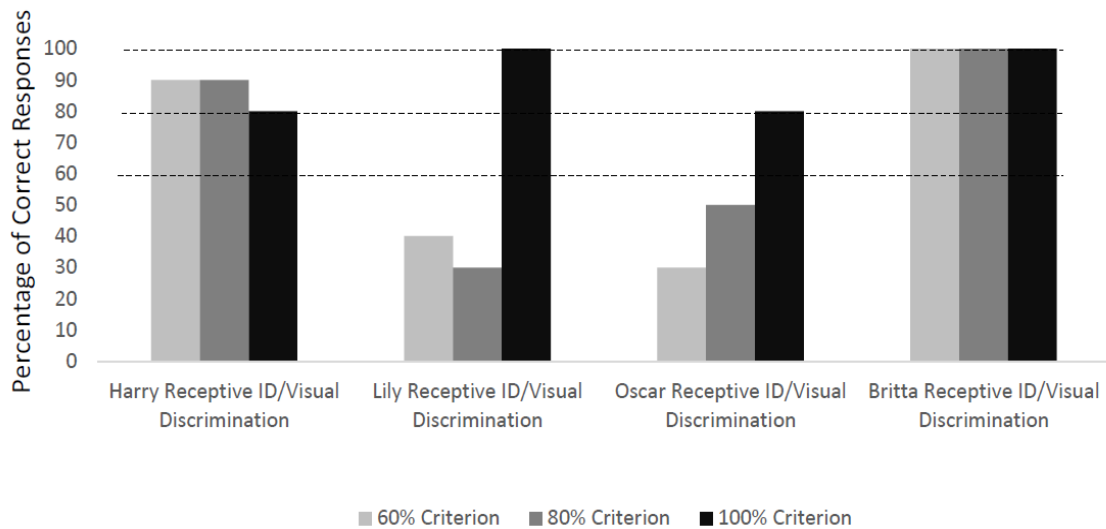
Appendix I

Experiment 2: Performance on First Maintenance Probe Session (10-trial block)

■ 60% Criterion   ■ 80% Criterion   ■ 100% Criterion
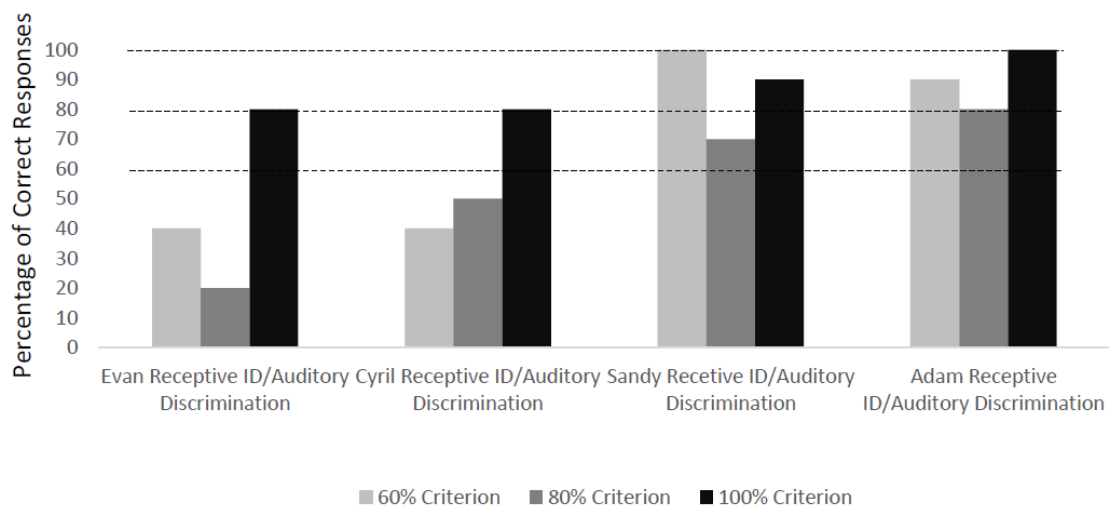
Appendix J

Experiment 3: Performance on First Maintenance Probe Session (10-trial block)

Appendix K

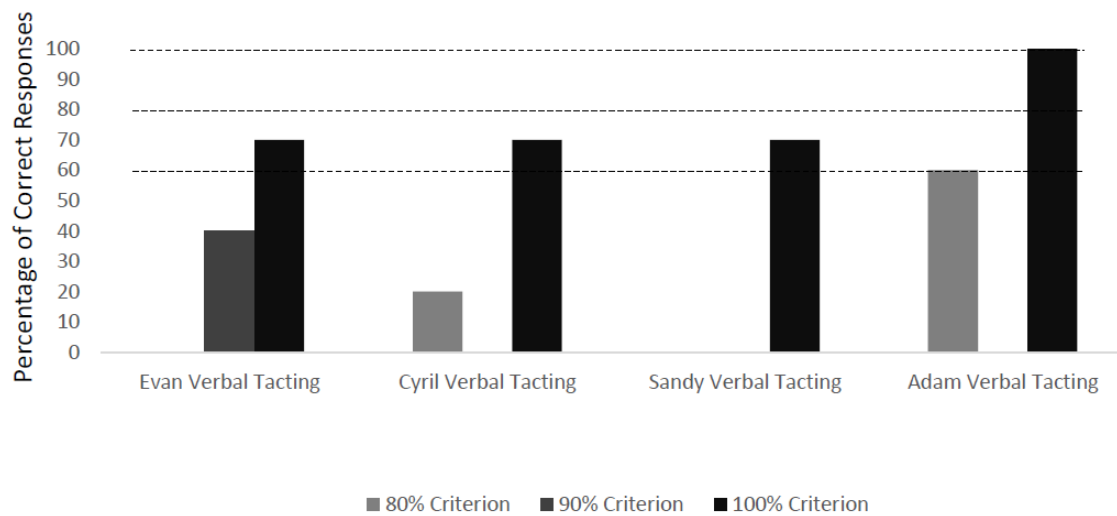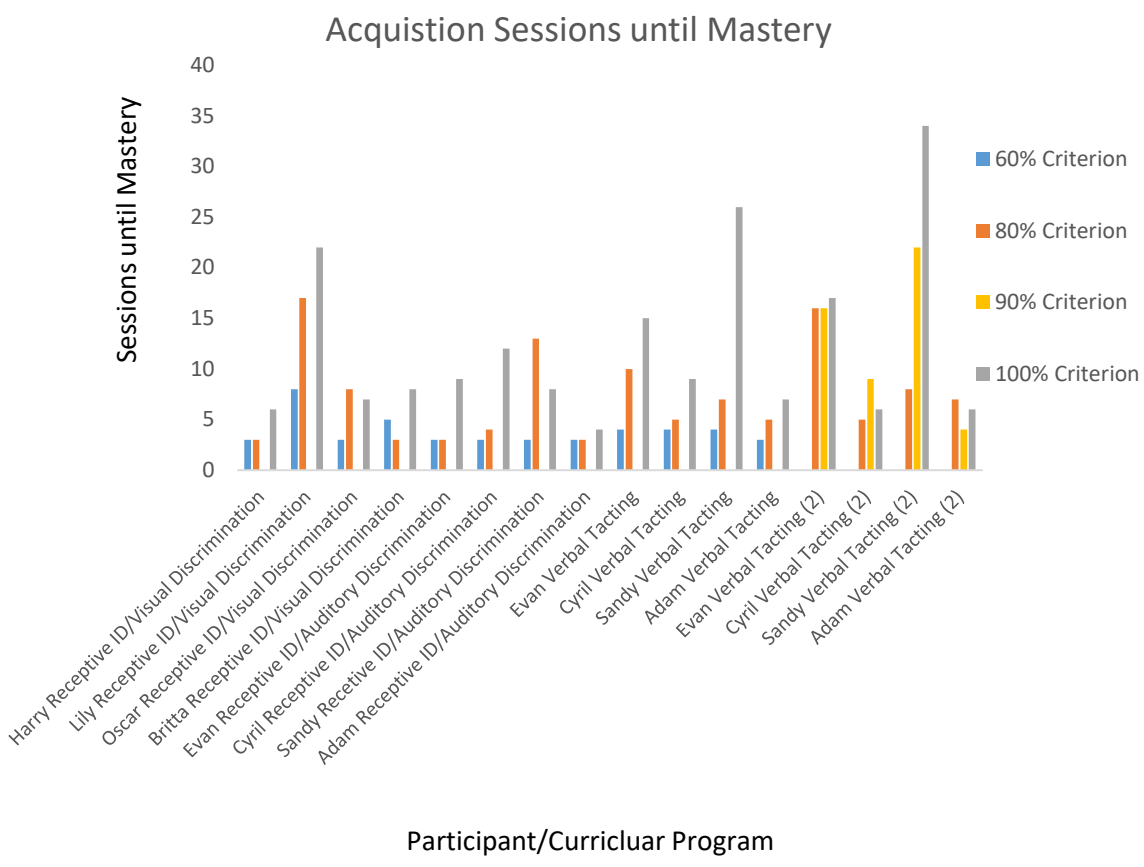Experiment 4: Performance on First Maintenance Probe Session (10-trial block)

Appendix L

Number of Acquisition Sessions until Meeting Mastery across Participants and Curricular Programs

Acquistion Sessions until Mastery

Appendix M

Scatter Plot Data of Performance on First Maintenance Probe as a Function Several Variables

Maintenance Performance/Sessions to Mastery

$y = -0.0249x + 60.005$
$R^2 = 3E-05$

Maintenance Performance/Criterion Level

$y = 0.604x + 9.9631$
$R^2 = 0.0771$

Maintenance Performance/Performance on Last Acquisition Session

$y = 0.887x - 17.82$
$R^2 = 0.1074$

Maintenance Performance/Average Performance on Last 3 Acquisition Sessions

$y = 1.0701x - 32.837$
$R^2 = 0.1453$