RESEARCH ARTICLE

A Preliminary Analysis of Mastery Criterion Level: Effects on Response Maintenance

Jessica L. Fuller^{1,2} · Daniel M. Fienup^{1,3}

Published online: 5 December 2017

© Association for Behavior Analysis International 2017

Abstract Educators use a mastery criterion to evaluate skill acquisition programming for children with autism and other developmental disabilities; however, to the best of our knowledge, there has been no research evaluating how the mastery criterion level of accuracy affects the maintenance of those responses. This study investigated the effects of 3 skill acquisition mastery criterion levels (50%, 80%, and 90% accuracy) on response maintenance. Following mastery of a set of skills, maintenance was evaluated once a week for 3 to 4 weeks. Three elementary school-age children diagnosed with autism participated. Overall, the outcomes suggested that higher mastery criterion levels (90% correct) produced higher levels of maintenance responding. Additional research in this area is needed to clarify how different parameters of mastery criterion affect the generality of skills.

Keywords Autism · Maintenance · Mastery criterion · Parametric analysis · Skill acquisition

- Educators can directly influence a learner's response maintenance by setting the mastery criterion at a high level, such as 90% correct.
- Lower criterion levels, such as 80% correct or 50% correct, produce lower levels of response maintenance.

- Researchers should investigate additional criterion levels, such as 100% correct.
- Researchers should evaluate additional components of mastery criterion, such as the number of consecutive sessions at the criterion level.

The use of a mastery criterion is ubiquitous in educational settings. Educators use a mastery criterion as a way to determine when a skill has been acquired. Once a participant's responding meets the predetermined accuracy level, then the educator may move to a less restrictive prompt level, choose novel targets to teach, or assess generalization or maintenance with the current target skill. There is an underlying assumption present in mastery criterion because behavior meeting this criterion signals the educator to do something else (e.g., provide less intrusive prompts, teach new target skills), which presumes some durability (i.e., maintenance) of that behavior when the current level of instruction is terminated. This should lead educators to the following questions: (a) What is mastery criterion? and (b) Is there research to suggest that behavior meeting mastery criterion is predictive of something else?

There is very little research supporting the efficacy of mastery criteria. This may be because educators and researchers treat mastery criterion and accuracy of responding as dependent variables. Most behavior analytic researchers incorporate mastery criteria in their studies to determine when to terminate a particular phase, but there is some discrepancy regarding the conditions under which a behavior is considered "mastered."

✉ Daniel M. Fienup
fienup@tc.columbia.edu

¹ Queens College, City University of New York,
New York City, NY, USA

² Manhattan Childrens Center, New York City, NY, USA

³ Department of Health & Behavior Studies, Teachers College,
Columbia University, 525 West 120th Street, Box 223,
New York, NY 10027, USA

What Are Mastery Criteria?

Accuracy-based mastery criteria can be conceptualized as containing at least two dimensions: level of performance and frequency of observations at that level. Level of performance can

be described as the number of correct responses compared to the total number of responses observed within a session. Educators and researchers typically set the criterion level as a percentage correct value between 80% and 100% accuracy. The frequency of observations at a level of performance refers to the number of consecutive sessions a particular level is observed. The mastery criteria reported in research articles vary across both dimensions. For example, studies have employed mastery criteria with 80% or 90% accuracy levels across two (e.g., Haq & Kodak, 2015; Majdalany, Wilder, Greif, Mathisen, & Saini, 2014) or three (e.g., Kocher, Howard, & Fienup, 2015) consecutive observations, and the researchers used criterion-level responding to signal the termination of a skill acquisition phase.

What Is the Evidence for Mastery Criteria?

To the best of our knowledge, no researchers have examined whether variations in the level of accuracy and number of observations affect the learning of children with disabilities who receive discrete trial instruction, despite the fact that mastery criteria are reported in nearly all skill acquisition publications in this area. The only empirical investigations of mastery criteria come from studies that manipulated criterion levels with typically developing adults who were learning college-level academic content (Carlson & Minke, 1975; Fienup & Brodsky, 2017; Johnston & O'Neill, 1973; Semb, 1974). These researchers found that higher criterion levels produced higher student performance. For example, Semb (1974) investigated the effects of manipulating mastery criterion and assignment length with two groups of college students enrolled in a personalized instruction course. The researcher examined high criterion (100% correct), low criterion (60% correct), and assignment length (a quiz based on one reading [short] vs. one cumulative quiz based on four readings [long]). Each participant experienced three conditions: high criterion on a short assignment, high criterion on a long assignment, and low criterion on a short assignment. Focusing on each participant's first attempt on a quiz, Semb (1974) found that a high mastery criterion for a short assignment produced better student performance (maintenance and generalization). Similar findings have been reported with fluency-based criteria (or criteria based on level of accuracy and time), and studies have shown that higher fluency-based criteria produce favorable educational outcomes (Dougherty & Johnston, 1996), especially with lower performing students (e.g., fourth- to seventh-grade students; Ivarie, 1986).

A Gap in Our Practice

Research is needed to evaluate the generality of the accuracy-based mastery criterion research findings from college

students to individuals with developmental disabilities engaged in individualized, discrete trial instruction. The findings regarding mastery criterion level with college students are promising and intuitive; however, the difference in educational needs between individuals with developmental disabilities and college students warrants research on this topic. As a first step toward empirically evaluating mastery criterion, this study parametrically manipulated the mastery criterion level of performance during skill acquisition and held the frequency of observations constant at one. Three specific criterion levels were evaluated: 50%, 80%, and 90%. We chose 80% and 90% criterion levels because these are commonly used in published research. We chose the 50% criterion level expecting that this condition would produce low maintenance. Researchers taught academic target behaviors to elementary school-age children diagnosed with autism to predetermined levels of performance and then assessed each set of skills once per week for 3 to 4 weeks.

Method

Participants, Setting, and Materials

Three male children diagnosed with autism participated. The study was conducted at the private special education school that the participants attended. Brandon was 5 years old and had attended the school for approximately two years. He communicated using an iPad and could independently mand for and tact a variety of items using this device. His academic programming included identifying sight words and using beginner-level math skills (e.g., number identification). Ryan and Jacob were 6- and 7-year-olds who had attended the school for 1 and 2 years, respectively. Ryan and Jacob both had large verbal repertoires and could vocally mand for and tact a variety of items. Each completed academic programming that included reading (e.g., sight words), beginner-level math (e.g., addition), and spelling. This study was approved by the university's institutional review board, and each participant's parent provided informed consent for the child to participate in this study.

The participants received one-on-one instruction throughout the school day, and all phases of the study were conducted at the participants' desks in their respective classrooms. The school used applied behavior analysis as its primary pedagogy, and instructors used a discrete trial training format for students' educational programming. In general, the school instructors set the skill acquisition mastery criterion at 90% correct across two consecutive sessions.

Several materials were used throughout the duration of the study, including toys (e.g., a toy cell phone), electronic games, and edibles that the participants had previously shown interest in. The materials for the target identification session and the

parametric analysis were different for each participant. The Dolch Second Grade Sight Words list (Dolch, 1948) was used for Brandon, the Dolch Third Grade Sight Words list (Dolch, 1948) was used for Ryan, and the Fry Third 100 Words list (Fry, 1994) was used for Jacob. Each participant used a token exchange system and received tokens following correct responses, which could be exchanged for preferred items on different ratio schedules of reinforcement (described later).

Response Measurement and Experimental Design

The researcher recorded response accuracy on each trial and reported the percentage of independent correct responses during a session. A correct response was defined as vocally stating (Ryan and Jacob) or physically spelling (Brandon [on an iPad]) the word that corresponded to the discriminative stimulus presented by the researcher. To be considered correct, the response had to begin within 3 s of the discriminative stimulus. An incorrect response was a response that did not correspond with the discriminative stimulus; if no response was initiated within 3 s of the discriminative stimulus, this was considered an incorrect response as well.

The researcher used an alternating treatments design to evaluate the effects of the parametric manipulation on response maintenance. The parametric analysis included three skill acquisition mastery criterion levels (50%, 80%, and 90%). The order of criterion-level conditions in a particular day was such that the researcher always began with a condition that was not conducted first on the prior day, and the remaining conditions were run in a different order than the previous day, thus approximating counterbalancing.

Procedure

The researcher conducted daily preference assessments with each participant. The researchers chose specific items based on a history of the participant showing interest in the item and the stimulus functioning as a reinforcer during skill acquisition. To begin each session, the researcher provided a participant with two stimuli and asked the participant to choose the specific stimulus for that session.

Target Identification The researcher identified target behaviors that were not in the participant's repertoire for use in subsequent phases. A participant sat at his desk across from the researcher. For a given trial, the researcher presented an instruction with any necessary stimuli and waited up to 3 s for a response. For Brandon, the researcher tested second-grade spelling words. The participant spelled each word on an iPad keyboard screen. To begin, the researcher placed the token board in front of him on his desk. The researcher presented the discriminative stimulus—"Spell [word]"—for each word. For Ryan and Jacob, the researcher tested second- and third-

grade sight words, respectively, presented on flash cards. The researcher held one flash card in front of the participant and vocally stated "Read the word." For all participants, the researcher provided praise plus a token contingent on a correct response, and each child earned five tokens to exchange for a preferred stimulus. The researcher provided no programmed consequence contingent on incorrect responding. The researcher evaluated each individual response two times and identified 15 total responses per participant that the participant responded to with 0% accuracy. During all sessions, the researcher praised appropriate sitting behavior. Additionally, for Brandon, the researcher interspersed mastered tasks during this phase (approximately every two trials).

Parametric Analysis of Mastery Criterion Level We randomly assigned five target behaviors identified in the previous phase to each of the three criterion-level conditions (50%, 80%, or 90%; see Table 1). During this process, we ensured that words with a similar number of letters were distributed equally across sets. The researcher spoke to the classroom teachers and ensured that these responses were not included in the participants' educational programming during the duration of this study.

Baseline, skill acquisition, and maintenance sessions were conducted in a similar manner to each other. The researcher administered 20 trials per session that included four trials of each of the five target responses presented in a randomized order. Each condition was run no more than once a day and approximately three to five days per week, except during maintenance when each condition was run one time per week. For each participant, the researcher provided praise and a

Table 1 Target behaviors

Participant	Type of program	Target responses per criterion level		
		50%	80%	90%
Brandon	Spelling	Today	From	Give
		Help	Went	Circle
		Jump	Saw	Walk
		Came	One	With
		Brown	They	Ate
Ryan	Sight words	Around	Because	Found
		Best	Call	Right
		Don't	These	Use
		Many	Fast	Sing
		Wish	Your	Why
Jacob	Sight words	High	Below	Believe
		Father	Small	Those
		Earth	Thought	Without
		While	Example	Enough
		Floor	Really	Seemed

token contingent on correct responding. During this phase, the researcher no longer interspersed mastered tasks for Brandon. For baseline, the researcher administered trials in a similar manner to trials in the target identification phase by providing the instruction and no assistance or prompts. During baseline, the researcher did provide praise and tokens contingent on correct responses. The researcher added an error correction procedure during the skill acquisition phase. Following an incorrect response, the researcher repeated the discriminative stimulus, provided a verbal model of the correct response for Ryan and Jacob and a textual model for Brandon, repeated the discriminative stimulus, and allowed 3 s for the participant to repeat the modeled response. The skill acquisition phase continued until the participant's accuracy level met or exceeded the respective mastery criterion level in one session.

The last phase of the parametric analysis began 1 week following the mastery of a respective set of targets. Maintenance sessions were identical to baseline sessions and were conducted once per week for 3 to 4 weeks. The timing of maintenance sessions was such that each set of targets was tested every 7 days following the specific day on which the targets were mastered during the skill acquisition phase.

Interobserver Agreement and Treatment Integrity

An independent observer collected interobserver agreement and treatment integrity data during 15%, 16%, and 21% of sessions for Brandon, Ryan, and Jacob, respectively. The researcher reviewed the target response and data collection procedures with the independent observer before each session. Interobserver agreement was calculated by dividing the number of trials in which both observers marked the same accuracy response (agreements) in a session by the total number of trials and multiplying that number by 100. Average percentage agreement was 93% (range of 90% to 100%) for Brandon and 100% for both Ryan and Jacob. For treatment integrity, the researcher developed a checklist that noted the procedural elements of the study. The steps included the researcher delivering the discriminative stimulus, delivering the reinforcer, and providing error correction to the participant contingent on incorrect responses. Treatment integrity was calculated by dividing the number of correct researcher responses in a session by the total number of responses and multiplying that number by 100. Treatment integrity was 99% (range of 98% to 100%) for Brandon and 100% for both Ryan and Jacob.

Results

Figure 1 displays participant data. During baseline, each participant responded with consistently low accuracy. When error correction was implemented for skill acquisition, each participant's performance steadily increased in accuracy until accuracy

met or exceeded the criterion level for the respective set of responses. To aid Brandon's acquisition, the researcher lowered the number of tokens required to exchange for preferred stimuli from five tokens to two tokens. Brandon met the 50%, 80%, and 90% criteria on the 9th, 10th, and 11th acquisition sessions by responding with 70%, 90%, and 90% accuracy, respectively. Ryan met the 50%, 80%, and 90% criterion on the 6th, 8th, and 9th acquisition sessions by responding with 70%, 80%, and 100% accuracy, respectively. Jacob met the 50%, 80%, and 90% criteria on the 7th, 10th, and 8th acquisition sessions by responding with 50%, 80%, and 90% accuracy, respectively.

After meeting the mastery criterion during the acquisition phase, the researcher began assessing maintenance once a week. For Brandon, there was overlap in maintenance outcomes. Brandon's accuracy with the targets assigned to the 50% criterion-level condition ($M = 68\%$, range of 60%–80%) were on an ascending trend and had similar accuracy to the 80% criterion-level condition ($M = 88\%$, range of 80%–100%) at the 4-week observations. Brandon responded with the highest maintenance to targets mastered at the 90% criterion level ($M = 93\%$, range of 90%–100%). Ryan responded 20% to 60% correct for targets that had been mastered at the 50% criterion level ($M = 35\%$), 50% to 70% correct for targets that had been mastered at the 80% criterion level ($M = 58\%$), and 70% to 100% correct for targets that had been mastered at the 90% criterion level ($M = 83\%$). Jacob responded 60% correct during all maintenance sessions for targets that had been mastered at the 50% and 80% criteria levels, and he responded 90% correct during all maintenance sessions for targets that had been mastered at the 90% criterion level. The researcher only collected 3 weeks of maintenance data for Jacob because the fourth week fell during a 2-week break from school.

For all three participants, the 90% criterion level consistently produced the highest level of maintenance responding 3 (Jacob) and 4 weeks (Brandon and Ryan) after the termination of skill acquisition. There was evidence of vertical separation of data paths during maintenance sessions for all three participants' data, suggesting a functional relation between the researcher-defined mastery criterion level during skill acquisition and response maintenance. However, the specific outcomes were somewhat variable across participants. For example, Ryan's responding during maintenance was very orderly, with each criterion level producing a different level of maintenance responding. Brandon and Jacob responded with higher levels of accuracy during maintenance to targets from the 90% criterion-level condition and responded similarly to targets from the 80% and 50% criterion-level conditions 4 weeks after skill acquisition was terminated.

Parametric Analysis

It is not surprising that different participants responded differently to mastery criteria (e.g., 50% and 80% criterion levels)

Fig. 1 The performances of Brandon (top panel), Ryan (middle panel), and Jacob (bottom panel) during baseline, skill acquisition, and maintenance

given their varying histories and specific repertoires. We performed a regression analysis to identify general patterns in response maintenance. Figure 2 displays the relationship between the researcher-defined mastery criterion during skill acquisition and accuracy during maintenance sessions. We collapsed all maintenance data for this analysis to increase the power to detect patterns. The y-axis displays the percentage

of correct responses during maintenance sessions. The x-axis displays the mastery criterion that the researcher used to determine when to terminate the skill acquisition phase for the set of target behaviors. Across all participants, the average maintenance performance was 53.6% (range of 20–80, $SD = 18.6$), 69.1% (range of 50–100, $SD = 16.4$), and 88.2% (range of 70–100, $SD = 7.5$) for targets mastered at the researcher-defined

Fig. 2 The scatterplot displays the relationship between the researcher-defined mastery criterion level (x-axis) and performance during maintenance (y-axis). Each gray circle represents an individual performance, and the darker-colored circles indicate overlapping data points. The black data points represent average performance based on the particular researcher-defined criterion level, and the dashed line represents the linear trend

criterion levels of 50%, 80%, and 90% correct or higher, respectively. The data in the graph show a pattern whereby lower criterion levels produced wider ranges of response maintenance with a lower level and higher criterion levels produced a smaller range of response maintenance with a higher level. Researcher-defined criterion levels significantly predicted maintenance performance, $b = .66$, $t(31) = 4.93$, $p < .001$. Researcher-defined criterion levels explained 44% of the variance in response maintenance, which was a significant proportion of the variance, $r^2 = .44$, $F(1, 31) = 24.34$, $p < .001$.

Discussion

What does it mean to have mastered a skill at a particular level of performance? The results of this study suggest that mastering a skill at a higher performance level predicts higher levels of maintenance 3 to 4 weeks after the termination of training. In this study, we manipulated the mastery criterion level and measured response maintenance. In the 50% criterion condition, participants needed to respond with accuracy between 50% and 100% correct during one session to terminate the skill acquisition phase because the criterion rule was 50% or higher. This resulted in participants meeting the criterion by responding with accuracy between 50% and 70% correct and maintenance between 20% and 80% correct at the end of the

study. In the 80% criterion condition, participants needed to respond with accuracy between 80% and 100% correct during training, and participants met the criterion by responding with 80% and 90% accuracy. Response maintenance varied across weeks (50% to 100%) and at the end of the study ranged from 20% to 80% accuracy. The least variable and highest level of response maintenance was observed in the 90% criterion condition. In this condition, participants needed to respond with 90% or 100% accuracy during training, and this resulted in response maintenance that varied from 70% to 100% accuracy, with all three participants responding with 90% accuracy at the end of the study.

The outcomes of this study are consistent with those of previous research that also found that higher mastery criteria resulted in higher levels of performance for college students (Fienup & Brodsky, 2017; Johnston & O'Neill, 1973; Semb, 1974) and typically developing school-age children (Ivarie, 1986). In this study, when researchers set the criterion to a higher level, participants responded with higher accuracy before training was terminated and maintained high levels of performance 3 to 4 weeks later. The outcomes of this study have strong implications for behavior analysts working with children with developmental disabilities, as this area of service delivery accounts for a large proportion of the work that behavior analysts conduct. A clinician should carefully consider the mastery criterion that he or she uses to indicate that a response has been mastered because this level of performance has predictive value for later behavior, such as the maintenance of that response. Across three participants with different individual characteristics and learning histories, final accuracy during skill acquisition accounted for 44% of the variability in maintenance scores, which makes this variable an important consideration when planning for behavior change that is durable over time. A promising aspect of this research is that optimizing a participant's performance through the manipulation of mastery criterion does not require a clinician to have any additional specialized skills beyond those needed to conduct discrete trial training.

A few components of mastery criterion warrant further study. The current study evaluated three criterion levels (50%, 80%, and 90%) and terminated skill acquisition following one observation of behavior that met or exceeded that level of accuracy. Two of three participants responded with similar accuracy during maintenance to targets mastered at 50% and 80% criterion levels. This was a surprising outcome. This could be interpreted to mean that the 50% criterion level produced better-than-expected maintenance performance or that the 80% criterion level produced worse-than-expected maintenance performance. We did not evaluate a 100% criterion level or levels below 50%, and these levels should be studied to gain a better understanding of the effects of the parameter of the mastery criterion level. Additional data may support a linear relationship between criterion level and

maintenance performance; however, it would not be surprising to learn that a range of low mastery criteria (e.g., 10% to 30% accuracy) produce 0% maintenance. It is unknown whether a 100% criterion level would produce similar or better maintenance performances when compared to those observed in this study in the 90% criterion condition.

Another component of mastery criterion that warrants further investigation is the parameter of the frequency of observations. In this study, we set the frequency of observations to one. Oftentimes, researchers and clinicians will specify a criterion level with a frequency of two (e.g., Haq & Kodak, 2015; Majdalany et al., 2014) or three (e.g., Kocher et al., 2015) consecutive observations of behavior at the particular level. We were surprised that target behaviors mastered at the 90% criterion level across one observation in this study were maintained at 90% correct at the end of the study for the three participants. We were surprised largely because we are accustomed to mastery criteria that incorporate frequencies of observations of two or more. In this study, we kept the frequency of observations to one so as to control for this variable while studying the criterion level. Future studies should parametrically analyze the frequency of observations to examine how this variable affects future responding. The primary dependent variable that may differentiate may be related to efficiency. At some point along the frequency of observations continuum, there will likely be no differences in effectiveness, but longer frequencies of observations may lengthen training time and result in less efficient skill acquisition and maintenance.

There were several limitations to the present study, all which potentially affect the generality of the outcomes reported herein. First, participants were taught different responses based on their existing repertoires and educational curricula. Teaching the same responses or sets of responses to participants may limit extraneous variables affecting mastery or maintenance of the skills. Second, we taught responses that were not in the participants' repertoire and randomly assigned same-grade target behaviors to conditions with attention to the length of words, but we did not control for other aspects of word difficulty. This process could have resulted in more difficult target responses being unevenly distributed across conditions, although the strong effects of mastery criterion suggest that this effect was minimal at best. Third, we taught responses using a specific error correction procedure. Many clinicians teach using prompt and prompt-fading procedures. We did not use a prompt-fading procedure, such as most-to-least prompting, because these procedures tend to produce accurate behavior at each prompt level, and this would have precluded us from stopping skill acquisition at an accuracy level of 50%, 80%, or 90% correct. The effectiveness of a teaching procedure may interact with the effectiveness of the mastery criterion, and the generality of our findings to teaching contexts with different teaching procedures remains unclear.

Recently, there have been several studies that have examined components of discrete trial instruction, such as the distribution of trials across sessions (Haq & Kodak, 2015), intertrial intervals (Majdalany et al., 2014), reinforcer magnitude (Paden & Kodak, 2015), the distributions of trials and reinforcers (Kocher et al., 2015; Ward-Horner, Cengher, Ross, & Fienup, 2017), and delays to reinforcement (Majdalany, Wilder, Smeltz, & Lipschultz, 2016). With no prior research evaluating the effects of mastery criterion with children with developmental disabilities, this study joins these studies in elucidating an important aspect of discrete trial instruction that affects skill acquisition outcomes. This study represents a first step in understanding the effects of mastery criterion level, and additional research could help to further clarify how different components of mastery criterion affect the maintenance and generalization of skills.

Compliance with Ethical Standards The research was approved by the institutional review board of the affiliated university. All parents of participants provided informed consent for the children to participate in this study.

Author Note The study was conducted by the first author in partial fulfillment of a master's degree in applied behavior analysis at Queens College, City University of New York. The authors wish to thank Julia Brodsky and Bryan Tyner for their helpful comments on the manuscript.

Conflict of Interest The authors declare that they have no conflicts of interest.

References

- Carlson, J., & Minke, K. (1975). Fixed and ascending criteria for unit mastery learning. *Journal of Educational Psychology*, 1, 96–101. <https://doi.org/10.1037/h0078676>
- Dolch, E. W. (1948). *Problems in reading*. Champaign, IL: Garrard.
- Dougherty, K. M., & Johnston, J. M. (1996). Overlearning, fluency, and automaticity. *The Behavior Analyst*, 19, 289–292.
- Fienup, D. M., & Brodsky, J. (2017). Effects of mastery criterion on the emergence of derived equivalence relations. *Journal of Applied Behavior Analysis*. <https://doi.org/10.1002/jaba.416>. Advance online publication.
- Fry, E. (1994). *1000 instant words: the most common words for teaching reading, writing and spelling*. Laguna Beach, CA: Laguna Beach Educational Books.
- Haq, S. S., & Kodak, T. (2015). Evaluating the effects of massed and distributed practice on acquisition and maintenance of facts and textual behavior with typically developing children. *Journal of Applied Behavior Analysis*, 48, 85–95. <https://doi.org/10.1002/jaba.178>
- Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Remedial and Special Education*, 7, 25–30.
- 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.

- Kocher, C. P., Howard, M. R., & Fienup, D. M. (2015). The effects of work-reinforcer schedules on skill acquisition for children with autism. *Behavior Modification, 39*, 600–621. <https://doi.org/10.1177/0145445515583246>
- Majdalany, L. M., Wilder, D. A., Greif, A., Mathisen, D., & Saini, V. (2014). Comparing massed-trial instruction, distributed-trial instruction, and task interspersal to teach tacts to children with autism spectrum disorders. *Journal of Applied Behavior Analysis, 47*, 657–662. <https://doi.org/10.1002/jaba.149>
- Majdalany, L. M., Wilder, D. A., Smeltz, L., & Lipschultz, J. (2016). The effects of brief delays to reinforcement on the acquisition of tacts in children with autism. *Journal of Applied Behavior Analysis, 49*, 411–415. <https://doi.org/10.1002/jaba.282>
- Paden, A. R., & Kodak, T. (2015). The effects of reinforcement magnitude on skill acquisition for children with autism. *Journal of Applied Behavior Analysis, 48*, 924–929. <https://doi.org/10.1002/jaba.239>
- Semb, G. (1974). The effects of mastery criteria and assignment length on college-student test performance. *Journal of Applied Behavior Analysis, 1*, 61–69.
- Ward-Horner, J. C., Cengher, M., Ross, R. K., & Fienup, D. M. (2017). Arranging response requirements and the distribution of reinforcers: a brief review of preference and performance outcomes. *Journal of Applied Behavior Analysis, 50*, 181–185. <https://doi.org/10.1002/jaba.350>