Mastery Criteria and the Maintenance of Skills in Children with Developmental Disabilities

Pitts, Laura; Hoerger, Marguerite

Behavioral Interventions

Published: 01/04/2021

Publisher's PDF, also known as Version of record

Received: 6 August 2020 | Revised: 3 February 2021 | Accepted: 4 February 2021

DOI: 10.1002/bin.1778

BRIEF REPORT

WILEY

Mastery criteria and the maintenance of skills in children with developmental disabilities

Laura Pitts | Marguerite L. Hoerger

School of Education, Bangor University, Bangor, UK

Correspondence
Laura Pitts, School of Education, Bangor University, Bangor, Gwynedd LL57 2AS, UK.
Email: edp72e@bangor.ac.uk

Abstract

Behavior analysts typically set a criterion for correct responding to determine when a skill is considered mastered. Practitioners often preset the criteria arbitrarily as there is little empirical evidence about the effects of differing mastery criteria on the maintenance of skills. The aim of the current study was to evaluate the effects of differing mastery criteria on skill maintenance. The impact of 80%, 90%, and 100% correct responding on the maintenance of sets of skills taught during discrete trial teaching was evaluated. Four children aged between 5- and 9-years-old with a diagnosis of Autism Spectrum Disorder participated in the research. Participants were taught to recognize words. Sets of words were taught until each of the three criteria were met, skill maintenance was probed after 1 week, and again once per week for another 3 weeks. Results demonstrated that skills were maintained at a response accuracy similar to the mastery criteria employed; with mastery criteria of 100% reliably producing higher levels of accurate responding during maintenance than a mastery criteria of 80% and 90%.

KEYWORDS
developmental disabilities, discrete trial teaching, mastery criteria, skill acquisition, skill maintenance

This study was completed in partial fulfilment of the first author's PhD degree in Education at Bangor University.

1 | INTRODUCTION

Educational programmes based on applied behavior analysis (ABA) for children with developmental disabilities and learning difficulties typically employ a range of teaching methods including discrete trial teaching (DTT) and natural environment teaching. In DTT teachers repeatedly present a stimulus, instruction, and prompt, and then reinforce correct responding until the child responds correctly and independently for a predetermined number or percentage of trials, often referred to as mastery criteria.

Mastery criteria are typically preset by the teacher or practitioner. Standards for mastery criteria vary across the literature and much of the existing evidence for the effects of differing mastery criteria involve evaluations which have been conducted with neuro-typical college students (Carlson & Minke, 1975; Fienup & Broadsky, 2017; Johnson & O'Neil, 1973; Keller, 1968; Semb, 1974). With limited empirical evidence, many practitioners have made recommendations for setting mastery criteria for individuals with developmental disabilities or learning difficulties based on clinical judgment and experience. Granpeesheh et al. (2014) recommend setting mastery criteria for individuals with Autism Spectrum Disorder (ASD) at 80% to 100% correct for two or three consecutive sessions. Leaf and McEachin (1999) recommend a general criterion for correct responding of 80% to 90% correct over 2 to 3 days with at least two different teachers; however, they suggest the criterion should be adapted to accommodate the individualized needs of learners.

Clinical practices regarding mastery criteria differ widely. Love et al. (2009) conducted a survey of clinical practices including mastery criteria practices with professional supervisors of early intensive behavioral intervention programmes; 62% (n = 119) of respondents reported using a percentage of correct trials across multiple sessions (e.g., 80% correct across two or three sessions), and 61% (n = 118) reported using a specific percentage of correct trials across two or more therapists (e.g., 80% correct across two or three sessions and at least two or more therapists).

Richling et al. (2019) conducted a survey of clinical practices related to mastery criteria practices with 194 board certified behavior analysts and doctoral level board certified behavior analysts. A majority (68%) of respondents reported using a specific percentage of correct trials. A further 28% of respondents set mastery criteria as a specific number of correct trials in a row, with a small minority (4%) using a rate of unit per time (fluency). The most frequently used mastery criterion was reported to be 80% accuracy over one or more consecutive sessions. While these studies provide an indication of what is commonly used, empirically validated evidence on what is effective is scarce. Little is known about the effects of differing mastery criteria on response maintenance. Response maintenance refers to the extent to which an individual is able to perform a skill after teaching is reduced or terminated. The number of successful, reinforced learning opportunities prior to the termination of teaching a skill is likely to effect response maintenance. A high number of successful learning opportunities leads to more frequent contact with positive reinforcement, which may strengthen behavior and result in behavior which is maintained more effectively. Thus, it would be reasonable to hypothesize higher mastery criteria would result in strengthened behavior, and behavior which ultimately maintains more effectively over time.

To our knowledge, only a few studies have been conducted to determine the effects of mastery criteria on response maintenance with children with ASD. Fuller and Fienup (2017) evaluated the effects of three different mastery criteria on response maintenance. Three children aged between 5- and 7-years-old with a diagnosis of ASD participated in the research. Literacy skills were taught until the children met different mastery criteria (50%, 80%, and 90%). Skill maintenance was evaluated once per week for 3 to 4 weeks. Results indicated the mastery criterion of 90% led to higher levels of accurate maintenance responding.

Richling et al. (2019) conducted an experimental evaluation of different mastery criteria with four children aged between 6- and 9-years-old with developmental disabilities. In two experiments discrimination tasks and vocal tacting were taught to 60%, 80%, and 100% mastery criteria. Maintenance probes were conducted weekly for 4 weeks after the participant achieved the mastery criteria. All of the participants demonstrated at least 80% accuracy in maintenance probes when skills had been taught to 100% mastery criteria for three sessions. When skills

were taught to 80% or lower accuracy, the participants all showed a significant decrease in accuracy during the maintenance probes. In a further experiment, vocal tacting tasks were taught to 80%, 90%, and 100% mastery criteria and response maintenance probes conducted following 1 week of criteria being met. Results revealed skills taught to 100% mastery criteria were maintained at the highest rate. These data suggest 80% mastery criteria may not be sufficient to promote the maintenance of skills in some children.

The current research evidence suggests higher mastery criteria result in skills which maintain more effectively over time. The research is limited to a small number of studies, which have looked at a limited number of skills and delays to maintenance probes. The current study seeks to replicate and extend these findings, which could influence how practitioners allocate teaching time to emerging skills. The differences between setting the mastery criteria at 80%, 90%, or 100% could have significant implications in terms of teaching time and skill acquisition rates. If the lower mastery criteria of 80% can establish mastery, and skills are demonstrated to maintain over time, then the overall time spent teaching may be lower and more skills can be taught in the same period of time. However, if lower criteria mean skills do not maintain effectively then time could be lost reteaching the skills later.

In a partial replication of the study by Richling et al. (2019) the current research compared the effects of three mastery criteria: 80, 90%, and 100% on accurate response maintenance. Three sets of stimuli within the same response class were taught to four children with ASD to evaluate the effects of the different mastery criteria on response maintenance. The stimuli sets were taught until responding met the predetermined mastery criteria, and response maintenance probe sessions began 1 week later. Maintenance probes were conducted once per week for 4 weeks.

2 | METHOD

2.1 | Participants and setting

Four children participated in this experiment. All of the participants had a diagnosis of ASD and a comorbid learning difficulty, two had an additional diagnosis of global developmental delay and sensory processing disorder. Inclusion criteria to be invited to participate in the research were for students to have basic receptive language skills, including being able to receptively identify items and pictures by touching, pointing, or gesturing with their hand.

Ruby was a 7-year-old female and Theo was a 5-year-old male; both could follow simple instructions and communicate using words and short sentences to mand, tact, and respond to intraverbals. At the time of the study Ruby had been participating in teaching which included DTT for approximately 3 years and Theo for a year. Theo had some low-level behaviors of concern (e.g., looking away, pushing task stimuli away with hands, throwing items with hands, and vocal protesting) related to escape from demand and attention. Noah was a 6-year-old male who was able to follow simple instructions and communicate using a Picture Exchange Communication System to mand and comment. He engaged in vocal protesting and hitting with his hands to access tangible items. Noah had been participating in teaching which included DTT for almost 2 years. Isla was a 9-year-old female who was able to follow simple instructions and communicate using a combination of sign language and vocals; she could use signs and words to mand, tact, and respond to intraverbals. Isla had been participating in teaching which included DTT for approximately 3 years.

The participants attended either a special educational needs school or a mainstream school where they received an education based on the principles of ABA. All participants were familiar with discrete trial teaching strategies, including all the reinforcement strategies, prompting and error correction procedures used in this study. Each session was conducted in a teaching room familiar to the participant. An ABA trained practitioner who was known to the participant conducted the sessions. The practitioner sat next to or opposite the participant during all sessions. An observer was present for some sessions for interobserver agreement and procedural integrity check

purposes. Sessions were conducted 1 to 3 times per day for 3–4 days per week based on participants and practitioners availability.

2.2 | Materials

Materials included tokens and token boards with space to attach five tokens; word cards (10.5 cm by 14.8 cm) on which words were printed in black text on a white background; data sheets; preferred edibles, preferred toys and items; an electronic timer to time access to preferred toys and items.

2.3 | Response definitions and measurement

Data were collected on participant responses during each discrete trial. Ten trials per session were run and the percentage of trials during which the participant responded correctly calculated. Correct responses were defined as the participant independently touching or pointing to the stimulus named by the trainer within 5 s of the instruction being given. Prompted responses were defined as trials during which prompts were provided to assist a correct response. Incorrect responses were defined as responses which did not correspond with the target stimuli or the participant not initiating a response within 5 s of the instruction being given.

2.4 | Interobserver agreement and procedural integrity

Interobserver agreement (IOA) was calculated by dividing the number of trials in which both observers agreed by recording the same correct responses in a session by the total number of trials and multiplying by 100. IOA was collected for 31.48% of baseline sessions, 37.80% of intervention sessions, and 30.13% of maintenance probe sessions. IOA was 99.36% for baseline sessions, 96.34% for intervention sessions, and 94.33% for maintenance probe sessions.

Procedural integrity checks were employed to ensure baseline, intervention, and maintenance probe sessions were implemented as designed. Observations of sessions were conducted and checklists were used to assess the correct implementation of procedures. Procedural integrity was collected on 26.67% of baseline sessions, 28.29% of intervention sessions, and on 25% of maintenance probe sessions. Procedural integrity was found to be 94.14% for baseline sessions, 96.87% for intervention sessions, and 97.22% for maintenance probe sessions.

2.5 | Target selection

Baseline sessions were preceded by target selection sessions. Targets consisted of written words which were selected if they were unfamiliar to the participant. A pool of potential target words were created with assistance from the participants teaching team. Probe sessions were conducted to ensure the proposed skills were unknown to the participants. During probe sessions, the teacher presented the written word in an array of 3 and gave the instruction “Touch… (target word).” There were no programmed consequences for participants responses during probe sessions. For each participant, nine unknown written words were split into three sets of three. Word length and the number of syllables per word were counterbalanced across sets, so words containing a similar number of letters and syllables were equally distributed to each criteria. One word from each set of three target words was taught until the participant responded independently and correctly to either 80%, 90%, or 100% mastery criteria.

Researchers ensured the target words were not included as teaching targets in participants regular education plan for the duration of the study.

2.6 | Design

A combined alternating treatment design and multiple-baseline design across participants was employed to evaluate skill maintenance following the teaching of sets of words to three mastery criteria; 80%, 90%, and 100%. Once mastery criteria had been met maintenance probes were conducted once per week for 4 weeks.

2.7 | Procedure

2.7.1 | Baseline

Baseline sessions mirrored the target selection trials. Ten trials were run during each baseline session. There were no programmed consequences for participants responses to the target stimuli during the baseline sessions.

2.7.2 | Intervention

Paired choice preference assessments were conducted at the start of each session to identify potential reinforcers to be used during teaching sessions. At the beginning of each teaching trial, the participant was presented with three cards which included the target word and two unknown distractors. The teachers said "Touch… (target word)." Least-to-most prompting procedures were used to facilitate learning. Least intrusive prompts were implemented initially, and prompts gradually increased as necessary: (1) Gestural (e.g., point prompts, hand gestures, nods of the head; (2) Positional (e.g., moving the target item closer to the participant); (3) Model (e.g., demonstration of the task); (4) Partial physical (e.g., gentle tap of the hand or arm); (5) Verbal prompt (e.g., repeating instructions, giving clues, explanations, verbal encouragement); (6) Full physical (e.g., hand over hand prompting).

Reinforcement in the form of a token was delivered following an accurate response on a fixed ratio 1 (FR1) schedule of reinforcement. Social praise (e.g., "well done," "nice work") was paired with the delivery of the token. The tokens were placed on a board, once five tokens were earned the participant exchanged them for a back-up reinforcer (e.g., a preferred edible, or 1 min of access to a preferred toy, item or activity). Following an incorrect response, the teacher quickly provided feedback such as "not quite" or "try again," reissued the instruction, and immediately prompted the correct response using the least intrusive appropriate prompt. The instruction was repeated to give the participant the opportunity to respond independently. Ten trials were run in each intervention session. Each correct response, prompted response and incorrect response were considered trials. Error correction procedures which followed incorrect responses were not counted as trials and therefore not recorded as such.

2.7.3 | Mastery criteria

The teaching sessions were terminated once the predetermined mastery were met criteria for three consecutive sessions. In the 80% and 90% conditions the first trial of each session was required to be correct; as if the initial trials were always prompted and subsequent trials were correct this wouldn't necessarily indicate true skill acquisition, and could have over inflated progress.

2.7.4 | Maintenance probes

Maintenance probe sessions were conducted once per week for four weeks after mastery criteria had been met. Ten trials per session were run. Maintenance probe sessions were identical to baseline sessions, with the exception that reinforcement was given following a correct response.

3 | RESULTS

Figure 1 displays the session data. Throughout baseline sessions each participants correct responses were relatively low. In the acquisition phase the participants accurate responding increased until mastery criteria were achieved or exceeded. During the maintenance phase Ruby demonstrated mean accurate responding of 77.5% (range 70%–90%) for skills taught to 80% mastery criteria, 95% (range 90%–100%) for skills taught to 90% criteria, and 98.33% (range 90%–100%) for skills taught to 100% criteria. For Theo a more distinct difference between mastery criteria and accurate maintenance responding was observed; 76.66% (range 60%–90%) for skills taught to 80% mastery criteria, 87.5% (range 80%–100%) for skills taught to the 90% criteria, and 98.33% (range 90%–100%) for skills taught to 100% criteria.

Noah's mean accurate responding for skills taught to 80% mastery criteria was 76% (range 60%–90%), for skills taught to 90% criteria; 90.83% (range 80%–100%), and 96.67% (range 90%–100%) for skills taught to 100%. Finally, Isla's mean accurate maintenance responding was 75% (range 60%–90%) for skills taught to 80% criteria, 92% (range 80%–90%) for skills taught to 90% criteria, and 98.33% (range 90–100) skills taught to 100% criteria.

Mean accurate maintenance responding across sets per mastery criteria was also analyzed (Table 1). Skills taught to 80% mastery criteria maintained at slightly lower levels lower than 80% with a mean of 76.46% (range 74.38–78.13) across sets, showing a small skill loss during maintenance when 80% mastery criteria was utilized. Mean accurate maintenance responding for skills taught to 90% mastery criteria was 91.25% (range 90.63–91.88), indicating no skill loss. Accurate maintenance responding for skills taught to 100% mastery criteria was 97. 92% (range 97.5–98.13), indicating a slight loss of 2.08%. The response accuracy during maintenance probes closely mirrored the response accuracy established during the teaching trials.

Maintenance responding data was simply acquired from probes and not sufficient enough to establish a functional relation, however, visual analysis of single case data displayed in figure one reveals some clear trends. All participants maintained accurate responding near or slightly below the levels observed during teaching sessions when the mastery criteria was set at 90 or 100%. A downward or variable tend is noted for some stimuli sets for all participants when the mastery criteria was set at 80% response accuracy.

4 | DISCUSSION

The purpose of this study was to compare the effects of three different mastery criteria on skill maintenance for four children with developmental disabilities. Practitioners may sometimes choose to teach to lower mastery criteria to allow more time to teach new skills; however, it is probable that skills mastered to correct responding on 80% of opportunities may not be sufficiently mastered to be robust over time. The most commonly used mastery criteria are reported to be 80% (Richling et al., 2019), however, skills taught to 80% mastery criteria have not been found to be sufficient in promoting skill maintenance in some children (Richling et al., 2019).

The findings of the current study add to an emerging evidence base suggesting higher mastery criteria results in skills which maintain more effectually. The present study evaluated maintenance with skills taught to 80%, 90%,

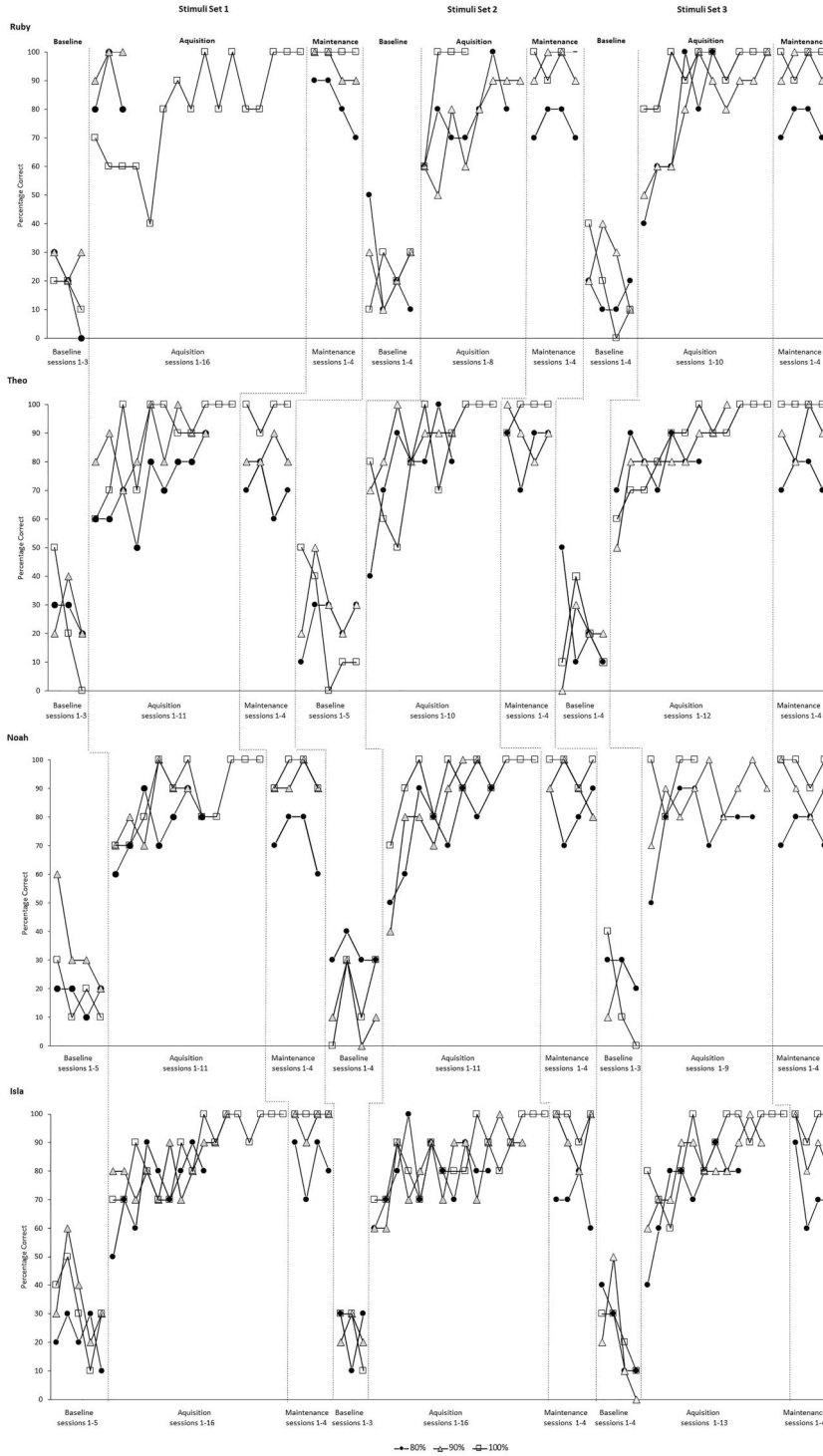

FIGURE 1 Percentage of correct responding during baseline, acquisition and maintenance sessions. Precisely 1 week elapsed between the final acquisition session and the first maintenance session. Subsequent maintenance sessions were conducted 1 week apart

TABLE 1 Mean percentage of accurate responding during maintenance trials across sets per mastery criteria

	Mean percent of accurate responding during maintenance trials		
Stimuli set	**80% mastery criteria**	**90% mastery criteria**	**100% mastery criteria**
Set 1	76.88%	91.25%	98.13%
Set 2	78.13%	91.88%	97.5%
Set 3	74.38%	90.63%	98.13%
Mean for all three sets	76.46%	91.25%	97.92%

and 100% mastery criteria. Skills taught to 100% mastery criteria resulted in the highest levels of maintenance following termination of teaching sessions, suggesting employing criteria in which 100% correct responding occurs for three sessions is more effectual for promoting the maintenance of skills. These data are not unexpected. Across participants and stimuli, we found the participants response accuracy during maintenance trials closely resembled the mastery criteria set during teaching trials. Therefore, if most discrete trials are taught to 80% mastery criteria, students will continue to respond accurately up to 80% of the time. We do not see untaught improvements in accuracy during maintenance trials. Skills that are taught to 100% mastery criteria are recalled more accurately than those taught to 80% accuracy. The pupils in this study showed a very small decrease in response accuracy when the skills were taught to 80% accuracy; these data contrast those of Richling et al. (2019) who noted a more significant loss of skills when mastery criteria was set at 80%.

Stimuli sets taught to 100% mastery criteria condition typically required a higher number of consecutive teaching sessions to reach criteria than those in the 80% and 90% criteria conditions. Stimuli sets in the 100% mastery criteria conditions consequently had the greatest number of learning opportunities, as well as the highest number of successful trials, and contacted the most positive reinforcement, all of which is likely to have strengthened skills and resulted in skills which maintained more effectively. Both the 90% and 100% mastery criteria conditions, however, resulted in maintenance levels above 90% accuracy, so in circumstances where an individual achieves 90% mastery criteria faster or over fewer teaching sessions than in the 100% mastery criteria then 90% could be an efficient criteria.

When setting mastery criteria practitioners need to consider the type of skills being taught and the level of accuracy the pupil will need for the skill to be generalized and useful in future. For skills related to health and safety it would be prudent to teach to 100% mastery criteria. For pivotal learning skills such as phonics, sight words, or basic arithmetic it may be advisable to teach to higher criteria as these skills are the foundation for future learning. This study did not include fluency measures, but it could be beneficial for future work to look at how both fluency and mastery criteria impact the maintenance of skills. Another point practitioners should consider when setting mastery criteria is functional or socially valid maintenance levels for the skills they are teaching.

This study is limited by the evaluation of only one category of skills: receptive identification of words. Words were assigned to conditions based on the number of letters and syllables they contained in attempt to control for some aspects of word difficulty. However, other aspects of word difficulty such as, number of phonemes, or diagraphs, trigraphs, or short or long vowel sounds were not accounted for specifically. Thus, in terms of difficulty there is a possibility that words could have been disproportionately allocated to their sets. However, it could be suggested that even if this were the case the impact of this was minor as results revealed fairly distinct differences in the effects of differing mastery criteria on skill maintenance. Future research should explore if and how mastery criteria affects the maintenance of other skills (such as tacting, intraverbals, visual performance tasks, etc.).

The current study evaluated accurate responding which occurred over three consecutive days. Dimensions of mastery criteria could warrant investigation, such as the number of consecutive sessions in which correct responses

occur (e.g., over one session, two consecutive sessions, or three consecutive sessions). Fuller and Fienup (2018) looked at correct responses which occurred only over one session. Increasing the mastery criteria over consecutive sessions at 100% correct would, however, result in increased successful learning opportunities and ultimately more contact with positive reinforcement; both of which may result in more durable behavior, and behavior which maintains more effectively. Setting criterion at 100% is not without its problems, it has been suggested a mastery criterion of 100% can lead to frustration and boredom with a task and may be unrealistic as mistakes can be due to variables other than lack of understanding (Leaf & McEachin, 1999). It is important to account for individual differences and make adjustments based on the individual's learning pattern.

Generalization across different stimuli, settings, and people wasn't programmed for in the current study but could be worth exploring. The process of generalizing across stimuli, settings and people would create additional practice opportunities which could consequently increase understanding, responding, and contact with reinforcement which could all subsequently influence skill maintenance.

We have investigated three different mastery criteria and their effect on skill maintenance in a population with which there is very little research into mastery criteria to date. In this study, we found most skills were maintained at a response accuracy similar to the mastery criteria used to decide when to terminate the direct teaching of a skill. The focus of future work should be on evaluating the specific components of mastery criteria itself across a range of different skills and populations. Future work should also look at other variables which may or may not be working in combination with, or potentially more heavily influencing skill maintenance such as teaching strategies, generalization procedures, and schedules of reinforcement.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ETHICS STATEMENT

The research was approved by the ethics committee of the affiliated university. All parents of participants provided informed consent for their children to participate in this study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author on reasonable request.

ORCID

Laura Pitts https://orcid.org/0000-0002-3943-1612
Marguerite L. Hoerger https://orcid.org/0000-0002-6564-2181

REFERENCES

Carlson, J. G., & Minke, K. A. (1975). Fixed and ascending criteria for unit mastery learning. *Journal of Educational Psychology*, *67*(1), 96–101. https://doi.org/10.1037/h0078676

Fienup, D. M., & Broadsky, J. (2017). Effects of mastery criterion on the emergence of derived equivalence relations. *Journal of Applied Behavior Analysis*, *40*, 843–848. https://doi.org/10.1002/jaba.416

Fuller, J. F., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion level: Effects on response maintenance. *Behavior Analysis Practice*, *11*, 1–8. https://doi.org/10.1007/s40617-017-0201-0

Granpeesheh, D., Tarbox, J., Najdowski, A. C., & Kornack, J. (2014). *Evidence based treatment for children with Autism: the card model*. Elsevier.

Johnson, J. M., & O'Neil, G. (1973). The analysis of performance criteria defining course grades as a determinant of college participants academic performance. *Journal of Applied Behavior Analysis*, *6*, 261–268. https://doi.org/10.1901/jaba.1973.6-261

Keller, F. S. (1968). Good-by teacher. *Journal of Applied Behavior Analysis*, *1*, 79–89. https://doi.org/10.1901/jaba.1968.1-79

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

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. https://doi.org/10.1016/jrasd.2008.08.008

Richling, S. M., Williams, W. L., & Carr, J. E. (2019). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis*, *52*, 107–117. https://doi.org/10.1002/jaba.580

Semb, G. (1974). The effects of mastery criteria and assignment length on college-participant test performance. *Journal of Applied Behavior Analysis*, *7*, 61–69. https://doi.org/10.1901/jaba.1974.7.61

How to cite this article: Pitts L, Hoerger ML. Mastery criteria and the maintenance of skills in children with developmental disabilities. *Behavioral Interventions.* 2021;36:522–531. https://doi.org/10.1002/bin.1778