



Mastery Criteria and Maintenance: a Descriptive Analysis of Applied Research Procedures

Cassidy B. McDougale¹ · Sarah M. Richling¹  · Emily B. Longino¹ · Soracha A. O'Rourke¹

Published online: 29 May 2019

© Association for Behavior Analysis International 2019

Abstract

Behavioral practitioners and researchers often define skill acquisition in terms of meeting specific mastery criteria. Only 2 studies have systematically evaluated the impact of any dimension of mastery criteria on skill maintenance. Recent survey data indicate practitioners often adopt lower criterion levels than are found to reliably produce maintenance. Data regarding the mastery criteria adopted by applied researchers are not currently available. This study provides a descriptive comparison of mastery criteria reported in behavior-analytic research with that utilized by practitioners. Results indicate researchers are more likely to adopt higher levels of accuracy across fewer observations, whereas practitioners are more likely to adopt lower levels of accuracy across more observations. Surprisingly, a large amount of research (a) lacks technological descriptions of the mastery criterion adopted and (b) does not include assessments of maintenance following acquisition. We discuss implications for interpretations within and across research studies.

Keywords Accuracy · Developmental disabilities · Maintenance · Mastery criteria · Skill acquisition

The field of applied behavior analysis (ABA) operates largely to develop skill acquisition and, in doing so, to promote maintenance of these skills (Baer, Wolf, & Risley, 1968). Practitioners often develop skill acquisition programs structured around meeting a predetermined mastery criterion as a means for promoting the maintenance of acquired skills (Luiselli, Russo, Christian, & Wilczynski, 2008). For example, a mastery criterion based on accuracy is associated with

several dimensions including the level of performance, such as a certain percentage correct, and the number of observations across which this level must be achieved, such as multiple sessions or days (Fuller & Fienup, 2018). Upon achieving “mastery,” a period of “maintenance” ensues, which may involve (a) no further teaching, (b) teaching less frequently, or (c) conducting maintenance probes at various intervals to determine if additional teaching is necessary. Practitioners assume that mastered skills will maintain both in the training context and be evoked in relevant situations. As such, it is assumed that achieving some level of mastery, in and of itself, is predictive of subsequent performance of that skill.

However, until recently, there has been a lack of research focused on identifying specific components of mastery criteria employed by researchers and practitioners and, consequently, evaluating the extent to which those practices lead to skill maintenance. Sayrs and Ghezzi (1997) conducted an analysis of the empirical articles reporting steady-state criteria published in the *Journal of Applied Behavior Analysis* (JABA) from 1968 to 1995. Results indicated that fewer than 20% of published articles describe a steady-state criterion, which could include statistical, mastery, or visual analysis criteria. Of the articles that reported steady-state criteria, the researchers identified an increasing trend in reported mastery criteria

Research Highlights

- Components of the teaching arrangement (e.g., the specific mastery criterion) are independent variables that impact learning outcomes such as skill maintenance and generalization.
- Applied researchers more commonly use higher percentages correct across fewer sessions to determine mastery, whereas clinicians use lower percentages correct across additional sessions.
- A large portion of the reviewed research lacks technological descriptions of the mastery criteria used and assessment of maintenance following acquisition.
- Further research is needed to determine what components of mastery criteria are important for promoting maintenance.

✉ Sarah M. Richling
smr0043@auburn.edu

¹ Department of Psychology, Auburn University, 226 Thach Hall, Auburn, AL 36849-5214, USA

approaching 100% in the most recently analyzed years. Researchers did not describe the variations of mastery criteria reported across empirical studies.

A more recent analysis conducted by Love, Carr, Almason, and Petursdottir (2009) sought to identify common practices of clinicians within the field of ABA. Researchers distributed a 43-question Internet survey to professional supervisors working in early intensive behavioral intervention programs with individuals diagnosed with autism. Researchers included several questions regarding strategies to promote skill acquisition, maintenance, and generalization. Results of the survey indicated the most common criteria for mastery generally required a certain percentage correct across multiple sessions (62% of reported) followed by a certain percentage correct across therapists (61% of reported). Data regarding more specific features of mastery criteria were not collected. Additionally, 98% of respondents reported including strategies to promote the maintenance and generalization of skills, which most often (50% of reported) consisted of reintroducing mastered targets in isolation or interspersed with other programs daily.

Richling, Williams, and Carr (in press) conducted a more comprehensive survey related to the mastery criteria adopted by Board Certified Behavior Analysts (BCBAs) and Board Certified Behavior Analysts at the doctoral level (BCBA-Ds) working as practitioners with individuals with intellectual disabilities (IDs). Similar to the findings of Love et al. (2009), most clinicians reported determining mastery as a certain percentage of correct trials across sessions. More specifically, the majority of clinicians (52%) reported using an 80% criterion across three sessions. Very few clinicians (7%) reported using a 100% mastery criterion in their practice.

Although mastery criteria are ubiquitous in clinical settings and in empirical skill-acquisition literature, there have only been two studies evaluating the relationship between various dimensions of mastery criteria and maintained responding with individuals diagnosed with autism and IDs. Although there is some literature evaluating mastery criteria directly, this research typically utilized undergraduate students as participants (e.g., Johnston & O'Neill, 1973; Semb, 1974). Thus, the extent to which findings apply to other populations is unknown. Recently, Richling et al. (in press) conducted an evaluation of maintenance following skill acquisition when applying differing mastery criteria (i.e., 80%, 90%, and 100% correct across three sessions) across skills with several individuals diagnosed with IDs. Results showed that a mastery criterion of 80% correct across three sessions was not sufficient to promote maintenance. Additionally, a criterion of 90% correct across three sessions did not produce consistent maintenance. By contrast, results showed that a criterion of 100% accuracy across three sessions was the most effective for promoting maintenance following skill acquisition.

Results from a similar study conducted by Fuller and Fienup (2018) support the findings that a higher mastery criterion may promote higher percentages of maintained responding following mastery. However, the authors found that a criterion of 90% across one session was effective in promoting maintenance. These results are in contrast with Richling et al.'s (in press) results regarding the 90% criterion. This difference could be due to several procedural differences. First, Fuller and Fienup (2018) implemented the 90% mastery criterion across only one session, but that session was composed of 20 learning trials. Richling et al.'s (in press) study utilized the same percentage across three sessions, but each session was composed of only 10 learning trials. In addition, it is possible the number of targets introduced in a massed-trial format may impact maintenance. Richling et al. (in press) introduced novel targets once individuals mastered each experimental target set, resulting in a greater number of targets being taught overall. By comparison, Fuller and Fienup (2018) did not report introducing any additional targets for teaching once participants achieved mastery for a target set.

Taken together, these studies suggest the mastery criteria adopted by practitioners (Fuller & Fienup, 2018; Richling et al., in press) may not promote skill maintenance. Several explanations for the use of procedures that do not have empirical support are possible. First, given the previous lack of research on this topic, one may argue that the procedures adopted by practitioners are based on lore rather than on empirical evidence. This is further supported by data indicating that the majority of practitioners report adopting specific mastery criteria that they observed during their supervisory training experience (Richling et al., in press).

However, a potential second explanation is that the research in this area is preliminary and incomplete. That is, further research may actually suggest an 80% criterion is sufficient once it is combined with various other components of mastery. As such, these practices adopted by practitioners may be reinforced by contingencies within their individual environments in the form of observed maintenance, rather than rules established by empirical reports.

A third possible explanation is that early supervisors in the field of ABA adopted specific mastery criteria they encountered in general behavior-analytic skill-acquisition research. Although this research did not directly manipulate mastery criteria as an independent variable and subsequently evaluate its effects on maintenance, specific mastery criteria may have been utilized in combination with other acquisition procedures and may have produced maintained responding. That is, if researchers previously (a) utilized certain mastery criteria within their research and (b) reported skill maintenance data, and (c) those criteria were consistently associated demonstrations of maintenance, there may be indirect empirical support for their use. No such evaluation of the mastery criteria reported in empirical research has been conducted to date.

Therefore, the purpose of this study is to systematically review recent applied behavior-analytic research to identify commonly used mastery criteria and the associated maintenance reported by investigators conducting skill-acquisition research. These results are directly compared to the common components of mastery criteria utilized by practitioners as reported by Richling et al. (in press).

Method

Data Collection

Authors included articles published between the years 2015 and 2017 for descriptive analysis. The analysis included articles from three peer-reviewed behavior-analytic journals that commonly publish interventions involving individuals with developmental disabilities: *JABA*, *Behavior Analysis in Practice* (BIP), and *Behavioral Interventions* (BIN). Authors searched each journal manually using a university search engine and evaluated articles published during each year for inclusion in the analysis. Inclusion criteria required that the article involve the implementation of a skill-acquisition intervention. *Skill acquisition* was defined as a peer-reviewed, published article for which the main purpose was to increase the frequency or accuracy of one or more behaviors. For inclusion, articles had to involve the direct manipulation of independent variables to alter some dimension of behavior. Descriptive analyses and interventions involving nonhumans were not included in the analysis.

Authors collected additional data across six dependent variables (listed next) for each article that met the aforementioned inclusion criteria. Coding within each of these categories was not mutually exclusive. For example, if an article reported that session-based criteria were used in Experiment 1 and trial-based criteria were used in Experiment 2, authors scored the article as both session based and trial based. As such, overall percentages of these categories total more than 100%. One author independently scored six dependent variables and included the following:

1. Utilization of specific mastery criteria method, including percentage of correct trials across sessions (i.e., session based), certain number of correct trials in a row (i.e., trial based), rate of response per a unit of time, and other (e.g., certain score across sessions, percentage across probes, duration across trials);
2. Specific percentage(s) utilized in the session-based criteria method;
3. Criteria for determining mastery across time (i.e., the required number of sessions, trials, trial blocks, or probes);

4. Whether authors programmed for or conducted probes to evaluate generalization (e.g., mastery or probes across settings, therapists, stimuli, target behaviors, adults);
5. Whether authors reported conducting maintenance at any time following the participant reaching specified mastery criteria; and
6. If maintenance was conducted, whether authors reported maintenance of skills following mastery (i.e., skills did maintain, did not maintain, were idiosyncratic across individuals, or unclear as described by the author).

Interobserver Agreement

To evaluate interobserver agreement (IOA), a secondary observer independently coded articles for each of the 3 years within each reviewed journal. For each year, the secondary observer scored an average of 32.6% (range 29.3%–34.6%) of articles reviewed for inclusion as skill acquisition. Thereafter, the secondary observer scored an average of 37.5% (range 30%–66.6%) of all articles identified as skill acquisition across the additional dimensions listed previously. Agreement was defined as identical information entered into the coding document and was calculated by dividing the number of agreements across each dimension by the total possible agreements for that dimension. IOA was 95% for identifying inclusion criteria across all articles. For articles identified as skill acquisition, IOA for each of the six recorded dimensions was 86%, 92%, 83%, 82%, 98%, and 97%, respectively.

Note that IOA scores below 90% agreement were obtained for the first, third, and fourth dependent variables. Inconsistencies in scoring these areas arose for several reasons. Within the reviewed literature, authors commonly described similar methods but referred to them inconsistently across articles (e.g., a group of trials being referred to as a *session* versus a *trial block*). Additionally, the dependent variables for determining the mastery criteria method and mastery criteria across time are interrelated. As such, an error in recording the criteria method typically resulted in an error recorded for the criteria across time. Errors regarding the scoring of generalization probes most often occurred when authors reported programming or evaluations of generalization but did not include the data sets.

Results

Skill-Acquisition Research Inclusion Criterion

Authors reviewed 473 articles for inclusion across the three specified journals; of these, 157 (33%) of the reviewed articles were identified as skill acquisition and were included in the subsequent analysis. The percentage of skill-acquisition

studies identified from JABA, BIP, and BIN were 39.6%, 24.5%, and 32.2%, respectively. There was an increasing trend in skill-acquisition publications across the 3 years reviewed for both BIP and BIN.

Mastery Criteria Method

Authors analyzed the 157 articles identified as skill-acquisition research to determine which mastery criteria method researchers utilized within the study. Table 1 compares the results of the current study with Richling et al.'s (in press) survey. Consistent with Richling et al.'s results, the most commonly utilized mastery criteria method was a certain percentage of correct trials (i.e., session based), which was utilized in 54% ($n = 84$) of skill-acquisition articles. This was higher than the reported utilization of various other mastery criteria (18%, $n = 29$), a certain number of correct trials in a row (i.e., trial based; 3%, $n = 5$), and rate of response per a unit of time (0.6%, $n = 1$). Researchers reported the use of a trial-based mastery criterion less often than practitioners in the Richling et al. study. Additionally, 26% ($n = 41$) of skill-acquisition articles did not report specific mastery criteria methods.

Specific Percentages Utilized During Session-Based Criteria

Authors included 84 articles utilizing session-based mastery criteria for analysis of specific percentages used across sessions to determine mastery. The top portion of Table 2 summarizes and compares these results with those of Richling et al. (in press). The most commonly used percentage for determining mastery across research studies was a 90% correct criterion (32%, $n = 27$). This criterion is higher than the 80% criterion most commonly reported by practitioners (Richling et al., in press). Twenty percent of research studies reported utilizing a 100% criterion ($n = 17$), higher than the percentage of practitioners using the same criterion (Richling et al., in press). Several studies reported a criterion between 81% and 89% (20%, $n = 17$) or an 80% criterion (18%, $n = 15$) during mastery across sessions. A small number of research articles also reported using a criterion between 91% and 99%

correct (8%, $n = 7$), whereas no articles reported utilizing a criterion below 80% for mastery.

Figure 1 shows the distribution of accuracy percentages reported by practitioners and research articles for session-based mastery criteria. The accuracy-level component of mastery criteria reported by practitioners is more heavily distributed between less stringent or lower percentages. In contrast, the accuracy-level component of mastery criteria reported within research more closely approximates a normal distribution.

Criteria for Determining Mastery Across Observations

Authors included articles identified as utilizing session-based or trial-based criteria or criteria of a certain percentage across consecutive probes for further analysis of criteria to determine mastery across observations. The bottom portion of Table 2 summarizes the results utilized during session-based criteria, the highest reportedly used mastery criterion, and compares them to those of Richling et al. (in press). The most commonly used criterion for articles reporting session-based mastery in research studies was requiring two sessions (60%, $n = 50$) at a specified percentage. Fewer articles reported requiring a certain percentage across one session only (21%, $n = 18$) or three sessions (21%, $n = 18$). This differs from clinicians, who most commonly reported utilizing a mastery criterion across three sessions (Richling et al., in press).

Figure 2 shows the distribution for the number of sessions required for mastery for practitioners and research articles. The distribution of the number of observations component of mastery criteria reported by practitioners approximates a normal distribution. This same component reported within research, however, is more heavily distributed toward less stringent criteria requiring fewer observations at a specific level of performance. This is in direct contrast with the distributions for the reported accuracy component of mastery criteria (Fig. 1).

Inclusion of Maintenance Probes

All articles identified as skill acquisition ($n = 157$) were included for further analysis to evaluate how often researchers

Table 1 Percentage utilizing specific mastery criteria

Criteria	Research studies (Total $N = 157$)	Richling et al. (in press) survey respondents (Total $N = 194$)
Certain percentage of correct trials (i.e., session based)	84 (54%)	132 (68%)
None reported	41 (26%)	N/A
Other	29 (18%)	N/A
Certain number of correct trials in a row (i.e., trial based)	5 (3%)	55 (28%)
Rate of response per a unit of time	1 (1%)	7 (4%)

Table 2 Percentage utilizing specific variables for session-based mastery criteria

Variable	Research studies (Total <i>N</i> = 84)	Richling et al. (in press) practitioners (Total <i>N</i> = 130)
Percentage for mastery		
100%	17 (20%)	9 (7%)
Between 91% and 99%	7 (8%)	6 (5%)
90%	27 (32%)	36 (28%)
Between 81% and 89%	17 (20%)	8 (6%)
80%	15 (18%)	70 (54%)
Below 80%	0 (0%)	1 (1%)
Sessions for mastery		
One	18 (21%)	10 (8%)
Two	50 (60%)	28 (22%)
Three	18 (21%)	65 (50%)
Four	1 (1%)	10 (8%)
More than four	1 (1%)	17 (13%)

reported conducting maintenance probes following mastery. The top portion of Table 3 summarizes these results. Forty-one percent (*n* = 64) of the studies reported conducting maintenance probes at some point following mastery of the target skill. Conversely, 58% (*n* = 91) of the reviewed articles did not report any conduction of maintenance following mastery. For the remaining 1% (*n* = 2) of the studies, it was unclear if authors conducted maintenance probes following skill acquisition.

Reported Maintenance of Skills

The current authors conducted further analysis of the 64 articles that included maintenance probes and corresponding results of those probes. The top portion of Table 3 summarizes these results. Of those articles that included maintenance probes, 61% (*n* = 39) reported successful maintenance of the target skill at some point following mastery. Thirty-three percent (*n* = 21) of the reviewed

articles reported idiosyncratic maintenance across participants, and 3% (*n* = 2) reported no maintenance of the skill. For 3% (*n* = 2) of the reviewed studies, no clear statements regarding the maintenance of skills during maintenance probes were included in the text.

Generalization

The 157 articles identified as skill acquisition were included for further analysis of reported generalization. The bottom portion of Table 3 summarizes these results. Fifty-nine percent (*n* = 92) of authors reported no specific programming for or probes to evaluate generalization following mastery. The most commonly reported generalization variable was programming for or conducting probes evaluating the skill across stimuli (17%, *n* = 27). This was higher than the number of those studies reporting conducting generalization across

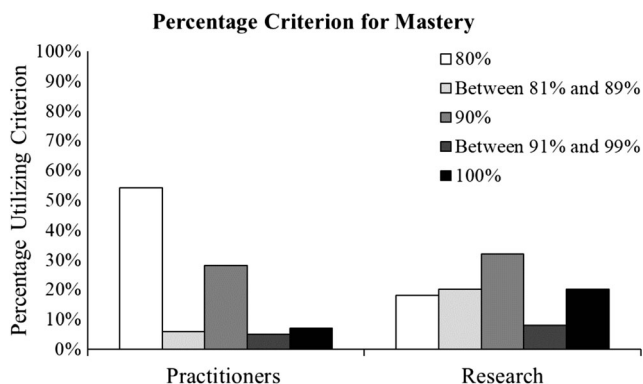


Fig. 1. Percentage of practitioners (Richling et al., in press) and research articles utilizing specific accuracy percentages for session-based criteria

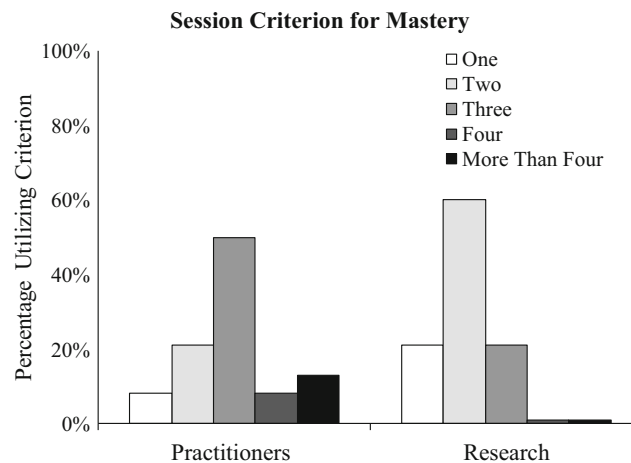


Fig. 2. Percentage of practitioners (Richling et al., in press) and research articles utilizing specific numbers of sessions

Table 3 Percentage utilizing additional variables during skill acquisition

Variable	Research studies (Total <i>N</i> = 157)
Maintenance	
Maintenance probes conducted	64 (41%)
Reported maintenance of skill	39 (61%) ^a
Idiosyncratic maintenance of skill	21 (33%) ^a
Unclear report of maintenance	2 (3%) ^a
No maintenance of skill	2 (3%) ^a
Generalization	
None	92 (59%)
Other	30 (19%)
Across stimuli	27 (17%)
Across environments	15 (10%)
Across therapists	10 (6%)
Across behaviors	8 (5%)
Across clients	8 (5%)

^a *n*/64

environments (10%, *n* = 15), across therapists (6%, *n* = 10), across behaviors (5%, *n* = 8), and across clients (5%, *n* = 8).

Discussion

In order to determine the variations of mastery criteria commonly utilized in behavior-analytic research, the current study analyzed recent skill-acquisition publications in several behavior-analytic journals. Authors compared these data to data obtained from a recent survey investigating the variations of mastery criteria commonly used in applied settings by BCBAs and BCBA-Ds (Richling et al., [in press](#)).

Overall, results indicate some overlap between the type of mastery criteria utilized by researchers and practitioners (i.e., session based, trial based, or rate of response per unit of time). That is, the majority of both researchers and clinicians report using an accuracy percentage to determine mastery. Regarding the various dimensions of accuracy-based mastery criteria, however, there are some notable differences between researchers and practitioners. With respect to the specific performance accuracy levels, results suggest that researchers are requiring higher levels, with 90% accuracy being the most common in research and 80% accuracy being the most common in clinical application (Richling et al., [in press](#)). However, results for the accuracy levels required by researchers appear to adhere to a relatively uniform distribution (Fig. 1), with less differentiation between the various accuracy levels. Results for the accuracy levels required by clinicians (Richling et al., [in press](#)), on the other hand, display a heavier distribution toward lower percentages (Fig. 1) with greater

differentiation between those using an 80% accuracy criterion versus other accuracy levels.

With respect to the specific number of sessions across which individuals apply these accuracies, opposite patterns are observed. That is, results for the number of sessions required by clinicians (Richling et al., [in press](#)) appear to adhere to a normal distribution (Fig. 2), with three sessions being the most common. Data on the number of sessions required by researchers, on the other hand, display a heavier distribution toward fewer sessions (Fig. 2), with two sessions being the most common.

There are implications of the findings that are worth noting. First, although the session-based mastery criterion was consistently the most commonly used across researchers (based on the current descriptive analysis) and practitioners (based on the findings in Richling et al., [in press](#)), the amount of skill-acquisition research articles that did not include a description of how mastery was determined (26%) is somewhat concerning. Although these data suggest an increase in reporting steady-state criteria when compared to the results of Sayers and Ghezzi (1997), it threatens one of the seven dimensions of ABA as described by Baer et al. (1968). One of the dimensions of ABA stipulates that behavior-analytic research is technological, meaning that researchers should be explicit in describing the procedures utilized. The failure to report how mastery was determined across recent behavior-analytic research may prove the replication of original findings experimentally or clinically by additional researchers or practitioners difficult. Additionally, failure to report mastery criteria procedures potentially restricts clinicians from arranging successful interventions for implementation and may limit the applicability of the research for practitioners.

Second, as previously mentioned, clinicians may be more likely to require lower accuracy levels, whereas researchers are more likely to require fewer sessions to determine mastery. Higher accuracy requirements utilized by researchers may be promising given the results of experimental evaluations conducted by Richling et al. ([in press](#)) and Fuller and Fienup (2018), suggesting that higher accuracy mastery criteria may be more effective in promoting maintenance. Those studies reporting that maintenance probes were conducted were also more likely to report maintenance of skills than to report no maintenance or idiosyncratic maintenance across participants. However, over half of the total research evaluated did not include generalization or maintenance probes.

Although the current results indicate that researchers are using higher accuracy criteria, they are requiring fewer sessions at those percentages to determine mastery. Currently, the literature has only evaluated these criteria across three sessions (Richling et al., [in press](#)) and one session (Fuller & Fienup, 2018), each using a different number of trials per session. In addition, it is worth noting that any mastery criterion that requires less than three sessions is not compatible

with the typical visual analysis standards for determining stability. The disparity between researchers' and clinicians' common practices may be a product of various treatment goals. Clinicians may be more frequently targeting skill-acquisition goals as described on an individualized education plan as determined by school staff and other professionals. Clinicians may also be under pressure to produce results efficiently, which may account for a more common acceptance of lower levels of accuracy. Researchers, however, may be offered more flexibility and control over the development and implementation of programming without as many time constraints or pressures from external sources. This could be one explanation for the higher levels of accuracy reported in research.

Further research is needed to determine how accuracy levels, number of observations (sessions), number of trials, and required time between sessions, as outlined by specific mastery criteria, may affect the maintenance of skills. For example, a future analysis could evaluate the extent to which the number of sessions required at a fixed accuracy percentage affects the maintenance of skills. Specifically, research should evaluate the contribution of various acquisition procedures as independent variables to the success of the entire teaching context for producing maintained responding. The effects of introducing additional training sets once individuals achieve mastery for a target set (such as those used by Richling et al., [in press](#)) should be evaluated, as this more closely resembles common clinical practices.

Third, over half of the research articles reviewed in the current study did not include maintenance probe sessions following acquisition. This is an important aspect of intervention, as described by Stokes and Baer (1977). Failure to conduct or report maintenance in the literature may have several effects on the field of ABA. First, failure to report these results may alter the value of the intervention for clinicians. Clinically, practitioners typically develop interventions with the overall goal that the individual will be able to perform the skill at a later time. As such, the likelihood of skill maintenance may be directly relevant to what type of interventions clinicians select to use for training. Reporting levels of skill maintenance could provide further support for specific interventions associated with higher levels of skill maintenance. The failure to report maintenance makes it difficult to determine what type of mastery criterion is most successful in promoting maintenance. A current review of the literature to determine which specific mastery criteria lead to maintenance may be futile. Less than 50% of the studies include maintenance probes and, thus, indirect evaluations of the mastery criteria utilized in these studies are impossible. Future researchers should include a precise description of the mastery criterion used consistently within a study, as well as report on the maintenance of skills following meeting this criterion.

Fourth, results of the current study also suggest a potential disconnect between criteria used by researchers and clinicians

that is inconsistent with evidence-based practice. Whereas clinicians are more likely to report utilizing an 80% criterion (Richling et al., [in press](#)), researchers were reportedly more likely to use a 90% criterion to determine mastery. Additionally, a larger number of research articles report using a 100% mastery criterion. It is promising that few research studies (18%) are utilizing an 80% criterion, which has proved ineffective for promoting maintenance (Fuller & Fienup, 2018; Richling et al., [in press](#)). However, these findings suggest that clinicians are commonly using mastery criteria that are not empirically based. Clinicians may be reducing the percentage criterion based on the perceived skills of individual learners. It is also possible that practitioners are requiring lower accuracy levels and then rolling those skills into maintenance programming following reaching mastery. As such, they may be detecting when responses do not maintain and subsequently conducting additional training. Certainly, such a practice is functional and acceptable; however, it may be inefficient overall. That is, if higher accuracy levels of performance are required early on, the skill may maintain for longer periods of time, such that an increasing number of previously acquired targets do not need to be retaught as often. In addition, if the skills being taught are components of a more complex skill or chain, errors across targets are likely to compound. Future research should evaluate the efficiency and effectiveness of various practices related to clinical maintenance checks.

Finally, the current analysis included many studies that used mastery criteria that included either/or criteria. These either/or criteria typically required fewer sessions at a higher percentage or an increased number of sessions at a lower percentage. This suggests that researchers, and potentially clinicians, are equating a criterion of a higher percentage across fewer sessions with a lower percentage criterion across additional sessions. However, there is currently no research to suggest these criteria are equal.

There are several limitations of the current analysis that should be considered. One limitation of the current review is that authors only reviewed select empirical journals for analysis. Inclusion of additional journals may alter the findings from the current study, resulting in research being more or less consistent with practitioner behavior. However, the authors included three commonly referenced behavior-analytic journals with higher percentages of skill-acquisition articles published across the 3 target years. Several other journals were considered but were determined to include a smaller ratio of acquisition studies. Future research may expand on the current analysis to provide a more diverse sample of scholarly outlets.

Another limitation is that the current review focused on the 3 most recent years of behavior-analytic research. Future analyses could include additional years to evaluate the methods for determining mastery over a longer period of time. However,

for the current analysis to be directly comparable to the findings of Richling et al. (in press), the authors selected a period of time similar to that of the survey. The current analysis reviewed skill-acquisition articles published across all participant populations, as opposed to selecting out skill acquisition specifically targeting individuals with autism spectrum disorder or other IDs. However, there is currently no evidence suggesting mastery criteria would differ across populations. Future research should evaluate the extent to which behavioral principles concerning mastery differ across populations and if these differences may warrant further examination of mastery criteria for specific populations.

The current analysis includes an aggregation of data across a wide array of independent variables within the training context, which could be viewed as a limitation. For example, two different studies may have utilized a 90% criterion across two sessions, but one study defined a session as an opportunity to perform multiple steps of a task chain, whereas the other defined a session as 20 trials of a single receptive labeling task. Ultimately, the goal would be to determine if there is a mastery criterion rule that is so robust that it can be used effectively across all permutations of the learning context; however, such a goal may be lofty and time-consuming on an individual level. Future research may consider disaggregating the current data and conducting a parametric analysis that explores the possibility of one mastery value (e.g., 80%) having a different impact on response maintenance depending on a number of other variables across the number of sessions, format of sessions (e.g., massed versus distributed), prompting strategies, and length of sessions, as well as a number of other variations. Such an analysis would provide greater detail regarding the specific circumstances in which various criteria may or may not result in maintained responding. In addition, a parametric analysis of extant data sets may alleviate some burden on conducting individual experimental analyses evaluating systematic changes to one part of a complex learning context and determining the impact on response maintenance.

Finally, the current analysis utilizes data obtained from a survey that represents only a small sample of willing respondents among a large clinician population. These data are then compared to research published across three highly competitive journals in which many strong researchers may not actually publish. As such, the data may be skewed and may render an apples-to-oranges comparison. This does not, however, mean the comparison should not be made. Despite these limitations, this study highlights inconsistencies in the use of mastery criteria within behavior-analytic literature and by practitioners in the field and provides a framework for continued analysis on an important topic. The lack of research specifically evaluating mastery criteria as an independent variable within acquisition studies is inconsistent with evidence-based practice.

As discussed, Richling et al. (in press) reported that clinicians commonly determine mastery criteria based on their clinical training and information passed down from their supervisors. As such, the mastery criteria adopted by practitioners is, at least in part, a product of tradition. Currently, however, there are contingencies in place to promote the engagement of practitioners with current research. These include requirements to obtain continuing education units reviewing and applying current practices in the field. However, Richling, Rapp, Funk, and Moreno's (2014) study reviewing publication rates of presentations at the Association for Behavior Analysis International conference suggests many of these presentations do not result in publications and therefore may not represent evidence-based practice. Future research should evaluate effective systems for bridging the gap between contemporary research and potentially rigid clinical practices.

Overall, there is a distinct need for research that evaluates each component of the learning context, such as the specific mastery criterion, as an independent variable that impacts learning outcomes. Moreover, the entire arrangement of teaching components and the collective impact on acquisition, maintenance, and generalization must be evaluated systematically. Such a concerted effort toward research of this nature will help to ensure the use of evidence-based treatment packages over the piecing together of science and lore.

Author Note This study was conducted by the first author, under the supervision of the second author, in partial fulfillment of the requirements for the master's degree at Auburn University.

Compliance with Ethical Standards

Conflict of Interest Cassidy McDougale declares that she has no conflict of interest. Sarah Richling declares that she has no conflict of interest. Emily Longino declares that she has no conflict of interest. Soracha O'Rourke declares that she has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- 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. <https://doi.org/10.1901/jaba.1968.1-91>.
- Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion levels: Effects on response maintenance. *Behavior Analysis in Practice, 11*, 1–8. <https://doi.org/10.1007/s40617-017-0201-0>.
- 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. <https://doi.org/10.1901/jaba.1973.6-261>.
- Love, J. R., Carr, J. E., Almason, S. M., & Petursdottir, A. I. (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/j.rasd.2008.08.008>.

- 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.
- Richling, S. M., Rapp, J. T., Funk, J. A., & Moreno, V. (2014). Low publication rate of 2005 conference presentations: Implications for practitioners serving individuals with autism and intellectual disabilities. *Research in Developmental Disabilities, 35*, 2744–2750. <https://doi.org/10.1016/j.ridd.2014.07.023>.
- Richling, S. M., Williams, W. L., & Carr, J. E. (in press). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis*.
- Says, D. M., & Ghezzi, P. M. (1997). The steady-state strategy in applied behavior analysis. *The Experimental Analysis of Human Behavior Bulletin, 15*(2), 29–30.
- Semb, G. (1974). The effects of mastery criteria and assignment length on college-student test performance. *Journal of Applied Behavior Analysis, 7*, 61–69. <https://doi.org/10.1901/jaba.1974.7-61>.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349–367. <https://doi.org/10.1901/jaa.1977.10-349>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.