BRIDGING THE GAP: EVALUATING A FUSION OF PROCEDURES FOR CONCEPTUAL LEARNING

Catherine L. Williams, Claire C. St. Peter, Madeline J. Murphy

WEST VIRGINIA UNIVERSITY

Conceptual learning is demonstrated when a learner responds when new examples are presented (i.e., generalization) but not when new nonexamples are presented (i.e., discrimination). Gradually increasing the number of examples and nonexamples taught together (i.e., set-size expansion) promotes conceptual learning with nonhumans. Although set size impacts the speed of acquisition with humans, its effects on conceptual learning have not been evaluated. Therefore, the primary purpose of the current study was to compare acquisition and conceptual learning during two procedures: set-size expansion and single, full set-size. College students were taught two biological concepts, one using set-size expansion and the other with the full set of stimuli. Participants were given feedback on the accuracy of their responses during instruction and tested (with no feedback provided) to assess conceptual learning. There were no systematic differences in accuracy during instruction, duration of instruction, or conceptual learning between the full set and the set-size expansion procedures. However, accuracy during instruction did not reliably predict conceptual learning, demonstrating that conceptual learning must be measured rather than assumed.

Keywords: acquisition; conceptual learning; discrimination; generalization; nonexamples; set size

Conceptual learning is often a primary goal of instruction. Conceptual learning is demonstrated when responding does occur in the presence of untaught examples (i.e., generalization; e.g., Fleming & Levie, 1993; Mechner, 1965) but does not occur in the presence of untaught nonexamples (i.e., discrimination). For example, teachers providing instruction on identifying letters want their students to correctly identify each letter outside of the classroom (generalization), while also correctly telling the difference between letters regardless of the font, color, or location (discrimination). Instruction (i.e., differential reinforcement) can produce accurate responding on training stimuli but may not reliably produce conceptual learning. The form of instruction changes the likelihood of conceptual learning.

When designing instruction to produce conceptual learning, some considerations are the number of stimuli included in practice and the way they are introduced. The number of stimuli that must be taught to produce conceptual learning is unclear (e.g., Bodily et al., 2008; Layng, 2019; Lazarowski et al., 2019; Nakamura et al., 2009; Tiemann & Markle, 1985), though including more stimuli during practice should improve conceptual learning. However, the number of total stimuli needed (full set) may be impacted by the number of stimuli introduced at a time. The number of distinct stimuli presented across trials at any given point in the practice process is called the set size.

Smaller set-sizes result in more frequent presentations of the same stimulus relative to larger set-sizes (when holding other aspects of the teaching procedure constant). This relatively dense exposure to each stimulus could facilitate acquisition but could also result in memorization of the stimuli rather than conceptual learning (e.g., Nakamura et al., 2009; Tennyson et al., 1972). Additionally, using small set-sizes may

Author Note: Address correspondence to: Catherine Williams (email: williamscl@uncw.edu; phone: 910-962-3288), Department of Psychology, Wilmington, NC, USA. https://orcid.org/0000-0002-4265-2467

This study was completed in partial fulfillment of the requirements for a Bachelor's degree by the third author, under the supervision of the first and second authors. Summary files are publicly available at https://researchrepository.wvu.edu/. Session by session data files are available from the first author upon request. All authors contributed to the study conceptualization and design. The experimental program was created by the first author. Experimental sessions were conducted by the third author, under the first and second others' supervision. The third author contributed to earlier drafts of introduction, method, and results sections of the manuscript. The first author drafted the discussion. The first and second authors provided revisions to all sections and made substantial contributions to the final contacts of the manuscript.

Other Author Information: Claire C. St. Peter https://orcid.org/0000-0003-0725-3169

https://doi.org/10.17605/OSF.IO/TH8EF

not result in acquisition of the necessary discriminations between examples and nonexamples because fewer distinct examples and nonexamples are shown in close temporal proximity to each other (e.g., Kodak et al., 2020). In summary, larger set-sizes may produce more robust generalization but slower initial acquisition relative to smaller set-sizes.

Set-size changes during acquisition have been programmed in at least two ways. With nonhumans, researchers gradually added stimuli to the existing stimulus set (i.e., set-size expansion) to evaluate the size at which conceptual learning was demonstrated (Bodily et al., 2008; Lazarowski et al., 2019; Nakamura et al., 2009). For example, the first set might start with four stimuli, then four more are added to make eight, and so on. These experiments demonstrated that the way the set size changes can impact conceptual learning but starting with the full set was relatively more efficient. With humans, researchers programmed set sizes by diving the full set of stimuli into sets of a particular number of stimuli together and compared efficiency across different set sizes of tacts (Kodak et al., 2020; Vladescu et al., 2021). For example, the first set might start with four stimuli, and then the second set will also have four stimuli, but they are different stimuli than in the first set.

In addition to a different way of manipulating set size, two differences between the procedures used in human (Kodak et al., 2020; Vladescu et al., 2021) and nonhuman (Bodily et al., 2008; Lazarowski et al., 2019; Nakamura et al., 2009) research complicate direct comparisons across areas. First, the human research evaluated natural concepts whereas the nonhuman research evaluated abstract concepts. Natural and abstract concepts are distinguished based on the role of physical relations between examples and nonexamples of a concept. Natural concepts (e.g., Herrnstein et al., 1976) are defined based on shared physical features, such as perceptual features, whereas abstract concepts are defined based on a relation beyond perceptual features of the stimuli themselves (e.g., Bodily et al., 2008). It is not clear whether practice that facilitates learning concepts defined by relations other than perceptual features will also facilitate learning concepts defined by perceptual features. Second, Kodak et al. and Valdescu et al. only measured and compared acquisition, they did not measure accuracy on trials with novel stimuli (i.e., conceptual learning). However, the goals of discrete trial training (e.g., Lerman et al., 2016) include conceptual learning. Despite these differences between human and nonhuman research in this area, the set-size expansion procedure used with nonhumans may address the clinical need for procedures that facilitate conceptual learning in the human research.

The purpose of this study was to evaluate the efficiency and effectiveness of the set-size expansion procedure to teach concepts beyond tacts with humans. In addition, no previous investigation included measures of generalization and discrimination to novel stimuli following trial-based instruction with different set-sizes. Therefore, a secondary purpose of this investigation was to evaluate the extent to which accuracy during instruction predicted conceptual learning across full-set and set-expansion procedures.

METHOD

Participants

Participants were three undergraduate students recruited through an online system provided by the university. Each participant signed up for a single, 3-hr experimental appointment; only one participant could sign up for each appointment time. Participants earned extra credit in a psychology class based on the amount of time they spent in the session. Participants were white and 18 to 21 years old (M =19.67). Two participants identified as male, and one participant (P3) identified as female.

Setting and Apparatus

Sessions were conducted in a 4.1-m by 3-m room equipped with a one-way mirror that allowed for unobtrusive monitoring of the participant during the experiment. The experimental task was presented using a custom Visual Basic program on a Microsoft Surface 3 tablet with a touch screen (27.5-cm display; 2160-pixel x 1440-pixel resolution). Sounds were played using the tablet's native speakers set to the 50% volume setting. All experimental procedures were conducted using this program, so it was not necessary to collect data from a secondary observer.

General Trial Structure

At the start of each trial, a white, 116 by 42 pixel (approximately 3.0 by 1.1 cm) rectangle appeared with the word "START" printed on it in black, size 20, Microsoft Sans Serif font. When the participant touched the rectangle, it disappeared and four stimuli appeared, one in each corner of the screen. Each stimulus was 250 by 250 pixels (approximately 6.6 by 6.6 cm).

The stimuli remained on the screen until the participant touched one. When a stimulus was touched, all stimuli disappeared and written feedback appeared in the center of the screen in black, size 24, Microsoft Sans Serif font. The feedback text, which differed based on trial type and participant response, disappeared after 1 s and was followed by a 1-s intertrial interval, during which a blank, white screen was shown.

Each session was broken into blocks of 24 trials so that each stimulus in the full set was shown an equal number of times, and each example was located once in each of the four locations. Only changes in stimuli or type of feedback indicated transitions between blocks. Two kinds of blocks were presented: teaching blocks and probe blocks. During teaching blocks, positive feedback followed correct responses, and corrective feedback and a correction trial followed incorrect responses. Positive feedback (i.e., reinforcement) occurred when an example was selected and consisted of text that read "Right! + $0.04" and a concurrent trumpet fanfare that played for 1 s. During the fanfare, the total amount of hypothetical money earned so far in the experiment was displayed for 0.25 s, disappeared for 0.25 s, and reappeared for 0.5 s with $0.04 added to the total. Although participants earned course credit, not actual money, for their participation, hypothetical money/points have previously been shown to function as a reinforcer in human operant tasks (e.g., Weiner, 1962).

Corrective feedback occurred when a nonexample was selected and consisted of text that said "Wrong. - $0.04" and a concurrent buzzer sound that played for 1 s. During the buzzer, the total amount of hypothetical money earned so far in the experiment was displayed for 0.25 s, disappeared for 0.25 s, and reappeared for 0.5 s with $0.04 subtracted from the total. If the total was $0.00, it remained at $0.00 ($0.04 were not subtracted). After the 1-s intertrial interval, a correction trial occurred that contained the same four stimuli as the previous trial, but the locations of the stimuli were randomly reassigned. Correct and incorrect responses on correction trials were followed with positive or corrective feedback, respectively. Correction trials neither counted toward the 24 trials in a block nor were responses during correction trials included in the summary data analyses for that block, though the number of correction trials was recorded. There was no limit to the number of correction trials. A response cost procedure was used to motivate participants to select the correct response in as few attempts as possible.

During probe blocks, all responses were followed by neutral feedback. Neutral feedback consisted of text that read "Selected." No auditory stimulus was played, and the total amount of hypothetical money was not shown or changed.

Participants received a 5-min break approximately every hour. When the participant completed a probe block more than 55 minutes from the start of the session or the previous break, a screen appeared telling the participant that it was time for a break.

Experimental Stimuli

Each participant was taught two concepts: influenza cells and plant cells. These concepts were chosen because they each have three must-have features but are independent from one another. The three must-have features of an influenza cell were spike surface protein, strand non-segmented RNA, and circular membrane proteins. For a plant cell, the three must-have features were chloroplast, vacuoles, and a cell wall. Additionally, stimuli had three features (color scheme, location of a feature, and number of some features) that could vary without affecting whether a stimulus was an example or a nonexample.

When any must-have feature was removed from an example of a concept, the stimulus became a nonexample of that concept. For each concept, we created two different types of nonexamples: close-in nonexamples and far-out nonexamples. Close-in nonexamples were created by replacing one of three must-have features. There was an equal likelihood that any one of the three must-have features would be replaced. Far-out nonexamples were created by replacing all three must-have features. The total number of features in the stimulus remained constant. Figure 1 shows three stimuli for each concept: an example, a close-in nonexample (missing one must-have feature), and a far-out nonexample (missing all three must-have features). The name of each must-have feature present is printed below each stimulus.

A total of 144 stimuli were created (72 of each concept). Of these, 48 stimuli (24 of each concept) were used to teach the concept during teaching blocks, and 96 stimuli (48 of each concept) were used only during probe blocks to assess conceptual learning. The 24 teaching stimuli for each concept contained six examples and 18 close-in nonexamples. The 48 probe stimuli consisted of 12 examples, 18 close-in nonexamples, and 18 far-out nonexamples.

Experimental Design

Teaching blocks were arranged using an adapted alternating treatments design to measure response acquisition in two conditions: full-set and set-size expansion. Teaching blocks occurred in sets of two, one of each condition. The conditions are described in detail below. The order of teaching blocks was randomly selected following each probe block. Probe blocks occurred at the start of the experiment (as a baseline) and after every two teaching blocks.

Pre-Experimental Procedure

When participants arrived, the experimenter reviewed an informed-consent form with them. After consenting, participants completed a demographics form. Participants then left their belongings (including all electronics) on a side table, and the experimenter read the following instructions:

> You will be shown four images at a time. You will earn $.04 of hypothetical money each time you select a correct image. When you select an image, you may be told if your selection is right or wrong or you may just be told that your selection is completed. A "completed" does not tell you whether your answer is correct or incorrect, only that you have successfully selected a response. When you are done or it is time for a break, a message will appear on the center of the screen telling you to knock on the door behind you. Do you have any questions?

The experimenter answered any questions by repeating information given during the consenting process or instructions provided above and then left the room.

Experimental Procedure

Teaching blocks

Each trial in a teaching block contained four stimuli: one example and three close-in nonexamples. In the full-set condition, every teaching block included all 24 stimuli. In the set-size expansion condition, teaching blocks began with a set size of four stimuli (one example and three close-in nonexamples). Teaching with these stimuli continued until 22 of 24 responses were correct in one block. Once this criterion was met, four new stimuli were added to the original four, resulting in a set size of eight (two examples and six close-in nonexamples). This process continued until all 24 stimuli were included (or the appointment ended).

Probe blocks

Probe blocks consisted of 24 trials: 12 trials containing novel influenza-cell stimuli and 12 trials containing novel plant-cell stimuli. The participant was not told where a response was correct. For each concept, six trials contained close-in nonexamples (one must-have feature replaced), and the other six contained far-out nonexamples (all three must-have features replaced). The order of trials was randomized at the start of each probe block.

Mastery Criteria

Two teaching blocks followed each probe until one of three criteria was met during a probe: (1) at least 92% (22 out of 24) correct on three consecutive teaching blocks in both conditions (six total blocks) that included 24 stimuli (acquisition mastery; supported by Wong et al., 2022); (2) at least 83% (5 out of 6) correct on trials with close-in nonexamples and 83% (5 out of 6) on trials with far-out nonexamples in each condition on one probe block (concept-formation mastery); or (3) three hours passed since the start of the experiment.

Dependent Variables

There were four dependent variables: response accuracy, duration of teaching, number of teaching blocks, and number of correction trials. An accurate response was defined as touching the stimulus that was the example. Response accuracy was calculated separately for teaching trials (24 trials per block), close-in probe trials (six trials per probe block per condition), and far-out probe trials (six trials per probe block per condition) by dividing the number of trials with correct responses by the number of trials of that type. The duration of teaching in each condition

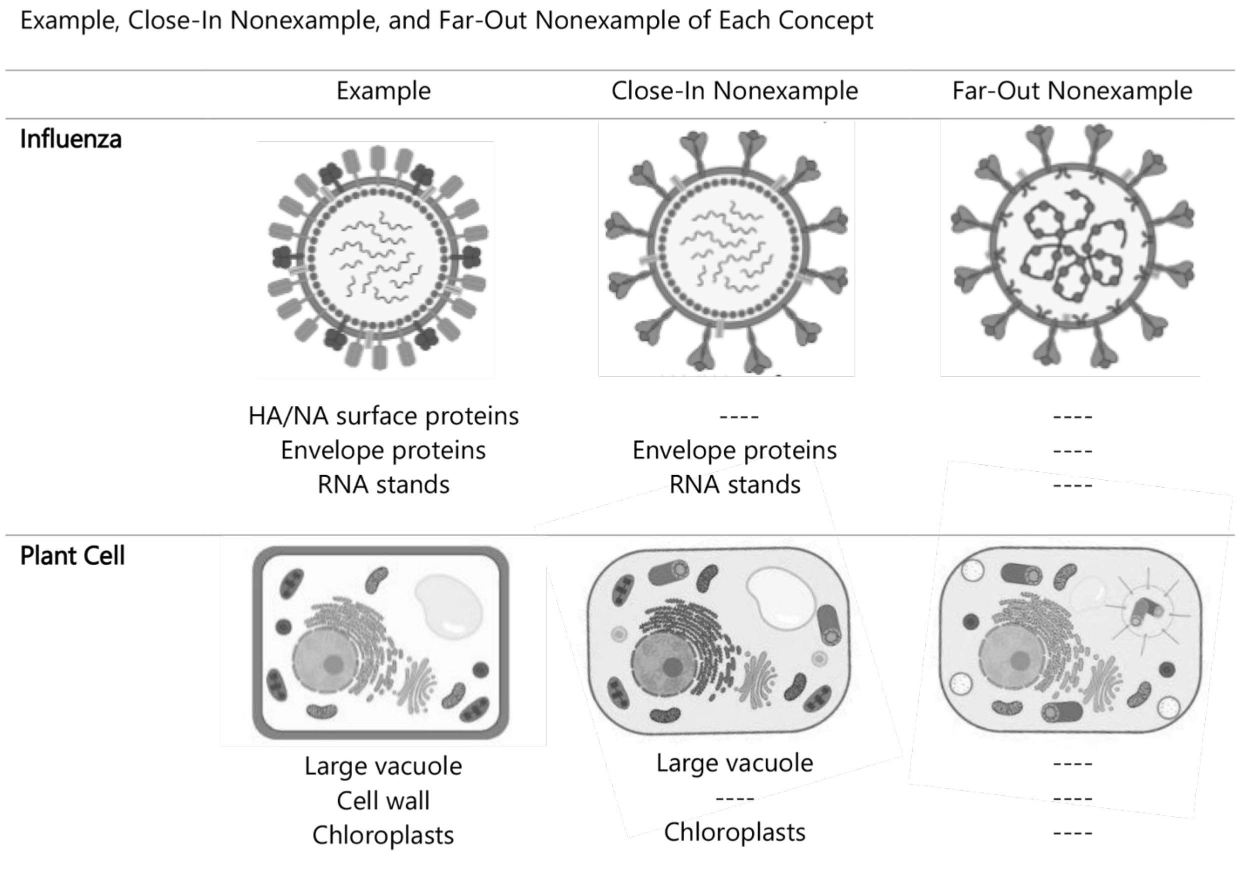

Figure 1. The names of the must-have features that are present are printed below the stimulus. Each row of text indicates the presense or absence of the must-have features of one concept. In the first column of stimuli, all must-have features are present, so the stimuli are examples of each concept. In the second column, one must-have feature of each concept is missing, so the stimuli are a close-in nonexamples of each concept. In the third column, all must-have features of each concept missing, so the stimuli are far-out nonexamples of each concept.

was calculated by summing the duration of each teaching block in that condition before the participant met the concept-formation mastery criterion (5/6, or 83%, correct across both close-in and far-out probe trials). For each condition, the number of teaching blocks to mastery was defined as the number of teaching blocks before the probe block when a mastery criterion was met. The number of correction trials to mastery in each condition was calculated by summing the number of correction trials in all teaching blocks before the block when each probe mastery criterion was met.

RESULTS

Figure 2 shows the percentage of correct responses on trials in the full-set condition (top graph) and set-expansion condition (bottom graph) during teaching and probe blocks for all participants. Accuracy was low during baseline probes. P1 (left panel) did not demonstrate acquisition or conceptual learning in either full-set or set-expansion conditions; the session ended after three hours. The similarity of examples and nonexamples seemed to influence performance; P1 responded accurately when trials included far-out nonexamples, but not when trials included close-in nonexamples (i.e., during both close-in probes and teaching trials).

The middle panel shows the results for P2. Like P1, accuracy was low during baseline and increased more rapidly in trials with far-out nonexamples than with close-in nonexamples. Unlike P1, set size appeared to influence learning. P2 responded more consistently and met the concept-formation mastery more rapidly in the full-set condition relative to the set-

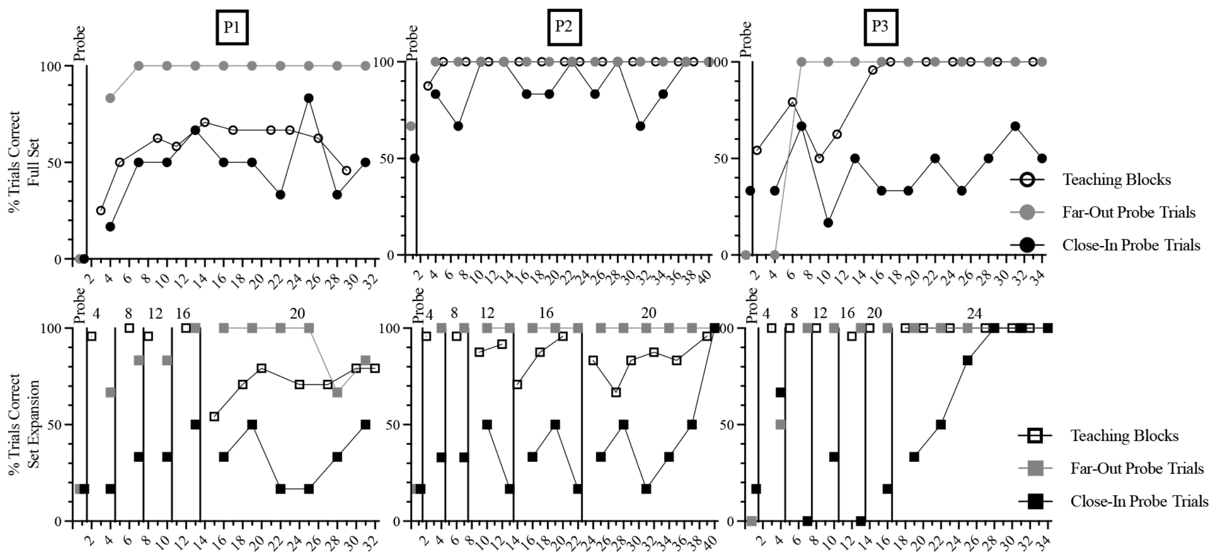

Figure 2. There are 24 trials per teaching block, six close-in trials per probe block, and six far-out trials per probe block. The data points before the first solid vertical line in both graphs are the baseline levels of accuracy on probe blocks. In the bottom graph, additional solid vertical lines indicate a change in the number of stimuli included in the teaching blocks in the set-expansion condition. The number of stimuli included is listed above the graph. All teaching blocks in the full-set condition (top graph) included 24 stimuli. P1 and P3 were taught the concept of "plant cells" in the set-expansion condition and "influenza cells" in the full-set condition. P2 was taught the concept of "influenza cells" in the set-expansion condition and "plant cells" in the full-set condition.

expansion condition. Notably, P2's accuracy in set expansion temporarily decreased each time new stimuli were added, suggesting P2 had memorized the previous stimuli rather than learning the concept.

The right panel shows accuracy for P3. Like P1 and P2, P3 had low accuracy during baseline and more rapidly learned to discriminate between examples and far-out nonexamples than close-in nonexamples. P3 responded more accurately during the set-expansion condition relative to the full-set condition and did not show predictable decrements in performance when new stimuli were added to the set. However, P3 did not respond accurately during probe trials with close-in nonexamples before the experiment ended (because P3 met the acquisition mastery criterion).

DISCUSSION

The primary purpose of this experiment was to address the need for efficacious trial-based procedures to teach concepts by testing a procedure used in the non-human literature to produce conceptual learning. We compared acquisition and conceptual learning when all examples and nonexamples were included at the start of teaching (full-set condition) to when the number of stimuli gradually increased (set-expansion condition). The set-expansion procedure had no consistent effect on acquisition or conceptual learning relative to using the full stimulus set.

We consider the utility of the set-expansion procedure to teach natural concepts to humans in the context of two measures of conceptual learning. Across both the set-expansion and full-set conditions, the relation between the probe accuracy and practice accuracy for each of these measures demonstrates a weak relation between practice accuracy and conceptual learning. The first kind of probe trials were those with examples and far-out nonexamples, which assessed conceptual learning when the nonexamples lacked all must-have features. Participants often met the mastery criterion in probe trials with far-out nonexamples before meeting the mastery criterion in teaching trials (which included only examples and close-in nonexamples with feedback on responding). The relation between accuracy during teaching and probes was weaker in the full-set procedure than in the set-expansion procedure. Therefore, practitioners should use caution when adopting set-expansion procedures to teach concepts. Initially beginning with a small set of stimuli could result in memorization rather than conceptual learning (e.g., Nakamura et al., 2009). In the current study, the number of teaching blocks required to produce accurate responding increased when the set size expanded, providing evidence that our participants memorized the small stimulus sets instead of attending to all the must-have features of the examples of the concept.

The second measure of conceptual learning was probe trials with close-in nonexamples, which assessed conceptual learning when the nonexamples lacked exactly one must-have feature. Accurate responding on these probe trials did not reliably occur. If it did occur, it was often several blocks after mastery on the practice blocks.

With the current data, the percentage correct during teaching blocks did not consistently predict conceptual learning. This finding underscores the importance of directly measuring conceptual learning when it is a desired outcome of trial-based teaching (e.g., DTT; Lerman et al., 2016). Researchers and practitioners should not assume that a certain number of stimuli, duration of teaching, or percent correct will necessarily produce conceptual learning. Instead, stimuli should be added, or teaching continued until conceptual learning is demonstrated (Stokes & Baer, 1977).

Two aspects of the present procedure may inform future research on set-size expansion procedures and conceptual learning. First, we incidentally measured the stability of accurate responding during probes for participants who met the mastery criterion in one condition before the other. Accuracy during far-out probe trials maintained for all participants in the set-expansion condition and for two of three participants in the full-set condition. In the set-expansion condition, accuracy during close-in probe trials maintained in the one case that maintenance could be evaluated (P3), but not in either case in the full-set condition (P1 and P2). These results provide preliminary evidence that maintenance might be less likely following instruction with the full set, although it is possible that participants were simply not motivated by earning hypothetical money. Future research should include a more robust reinforcer and more completely measure maintenance following initial mastery to evaluate the relation between teaching procedures, conceptual learning, and maintenance.

Second, future research could evaluate the set-size expansion and full-set procedures when different error-correction techniques were used. The current study included an error-correction procedure (re-present until independent; Carroll et al., 2015) used in previous studies involving trial-based learning (e.g., Kodak et al., 2020; Schnell et al., 2018, Vladescu et al., 2021). However, this error-correction procedure involves presenting multiple trials with the same four stimuli, which is essentially a temporary shift to a set size of four. Therefore, correction trials, when they occurred, could have facilitated learning on the full set by including a small sequence of trials with a smaller set-size. It is unclear if the use of this error-correction procedure enhanced the effectiveness of teaching with a full set. Future experiments could compare full-set and set-expansion procedures without this error-correction procedure.

A limitation of this study is that our baseline included only the probe stimuli and not the teaching stimuli. Because the concepts taught in this experiment were biological, participants may have had previous experience with these concepts. However, the close-in stimuli for teaching and probe trials were highly similar, and baseline accuracy on the probe stimuli was at or below 50% accuracy in all but one instance (P2, full-set condition). The higher initial baseline levels for this participant could have indicated previous knowledge that may have facilitated the rapid acquisition and conceptual learning observed in the full-set condition, so results should be viewed with some caution. Future research should also measure baseline responding with the teaching stimuli. The design could also be strengthened by embedding a multiple baseline or probe design across concepts.

More research is needed on trial-based instructional procedures that reliably produce conceptual learning. The current results suggest that neither the set-expansion nor the full-set procedures were sufficient to produce conceptual learning for all participants, even when responding was accurate during teaching. One improvement would be to include matched examples and nonexamples (e.g., Layng et al.,

2019) during practice. Matched examples and nonexamples contain the same can-have features so that only the must-have features differ. Until effective procedures that produce conceptual learning are identified, practitioners and researchers interested in conceptual learning should directly measure this outcome using generalization (or more precisely, concept-formation) tests that align with the goals of their instruction (e.g., discriminate between examples and close-in or far-out nonexamples). Practitioners interested in conceptual learning should report which procedures produce it and which do not (e.g., Fienup & Brodsky, 2017; Fuller & Fienup, 2018; Lee & Singer-Dudek, 2012; Richling et al., 2019).

REFERENCES

Bodily, K. D., Katz, J. S., & Wright, A. A. (2008). Matching-to-sample abstract-concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 34*(1), 178–184. https://doi.org/10.1037/0097-7403.23.1.178

Carroll, R. A., Joachim, B. T., St. Peter, C. C., & Robinson, N. (2015). A comparison of error-correction procedures on skill acquisition during discrete-trial instruction. Journal of Applied Behavior Analysis, 48(2), 257-273. https://doi.org/10.1002/jaba.205

Fleming, M. L., & Levie, W. H. (1993). Instructional message design: Principles from the behavioral and cognitive sciences. Educational Technology.

Fienup, D. M., & Brodsky, J. (2017). Effects of mastery criterion on the emergence of derived equivalence relations. Journal of Applied Behavior Analysis, 50(4), 843-848. https://doi.org/10.1002/jaba.416

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

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 2(4), 285-302. https://doi.org/10.1037/0097-7403.2.4.285

Kodak, T., Halbur, M., Bergmann, S., Costello, D. R., Benitez, B., Olsen, M., ... & Cliett, T. (2020). A comparison of stimulus set size on tact training for children with autism spectrum disorder. Journal of Applied Behavior Analysis, 53(1), 265-283. https://doi.org/10.1002/jaba.553

Layng, T. J. (2019). Tutorial: Understanding concepts: Implications for behavior analysts and educators. Perspectives on Behavior Science, 42(2), 345-363. http://doi.org//10.1007/s40614-018-00188-6

Lazarowski, L., Goodman, A., Galizio, M., & Bruce, K. (2019). Effects of set size on identity and oddity abstract-concept learning in rats. Animal Cognition, 22(5), 733-742. http://doi.org/10.1007/s10071-019-01270-5

Lee, G. T., & Singer-Dudek, J. (2012). Effects of fluency versus accuracy training on endurance and retention of assembly tasks by four adolescents with developmental disabilities. Journal of Behavioral Education, 21(1), 1-17. https://doi.org/10.1007/s10864-011-9142-9

Lerman, D. C., Valentino, A. L., LeBlanc, L.A. (2016). In R. Lang, T. B. Hancock, & N. N. Singh (Eds.), Early intervention for young children with Autism Spectrum Disorder (pp. 47-84). Springer. https://doi.org/10.1007/978-3-319-30925-5_3

Mechner, F. (1965). Science education and behaviour technology. In R. Glaser (Ed.), Teaching Machines and Programmed Learning II (pp. 441-507) National Education Association.

Nakamura, T., Wright, A. A., Katz, J. S., Bodily, K. D., & Sturz, B. R. (2009). Concept learning carryover effects from the initial training set in pigeons (Columba livia). Journal of Comparative Psychology, 123(1), 79. http://doi.org/10.1037/a0013126

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(3), 701-717. https://doi.org/10.1002/jaba.580

Schnell, L. K., Vladescu, J. C., Kodak, T., & Nottingham, C. L. (2018). Comparing procedures on the acquisition and generalization of tacts for children with autism spectrum disorder. Journal of Applied Behavior Analysis, 51(4), 769-783. https://doi.org/10.1002/jaba.480

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

Tennyson, R. D., Woolley, F. R., & Merrill, M. D. (1972). Exemplar and nonexamplar variables which produce correct concept classification behavior and specified classification errors. Journal of Educational Psychology, 63(2), 144-152. https://doi.org/10.1037/h0032368

Tiemann, P., & Markle, S. (1985). Analyzing instructional content: A guide to instruction and evaluation. Champaign. Stipes Publishing Co.

Vladescu, J. C., Gureghian, D., Goodwyn, L., & Campanaro, A. M. (2021). Comparing skill acquisition under different stimulus set sizes with adolescents with Autism Spectrum Disorder: A replication. Behavior Analysis in Practice, 14(1), 193-197. https://doi.org/10.1007/s40617-020-00506-y

Weiner, H. (1962). Some effects of response cost upon human operant behavior. Journal of the experimental analysis of behavior, 5(2), 201-208. https://doi.org/10.1901/jeab.1962.5-201

Wong, K. K., Fienup, D. M., Richling, S. M., Keen, A., & Mackay, K. (2022). Systematic review of acquisition mastery criteria and statistical analysis of associations with response maintenance and generalization. Behavioral Interventions, 37(4), 993-1012. https://doi.org/10.1002/bin.1885

Woolley, F. R., & Tennyson, R. D. (1972). Conceptual model of classification behavior. Educational Technology, 12(4), 37-39. https://www.jstor.org/stable/44417827