

*SHAPING ACADEMIC TASK ENGAGEMENT WITH  
PERCENTILE SCHEDULES*

ELIZABETH S. ATHENS AND TIMOTHY R. VOLLMER

UNIVERSITY OF FLORIDA

AND

CLAIRE C. ST. PETER PIPKIN

WEST VIRGINIA UNIVERSITY

The purpose of this study was to examine the use of percentile schedules as a method of quantifying the shaping procedure in an educational setting. We compared duration of task engagement during baseline measurements for 4 students to duration of task engagement during a percentile schedule. As a secondary purpose, we examined the influence on shaping of manipulations of the number of observations used to determine the criterion for reinforcement (the  $m$  parameter of the percentile formula). Results showed that the percentile formula was most effective when a relatively large  $m$  value (20 observations) was used.

DESCRIPTORS: shaping, academic engagement, percentile schedules

Shaping is a powerful method used to promote changes in existing behavioral repertoires. *Shaping*, also known as the method of successive approximations, can be defined as the gradual modification of some property of responding by differentially reinforcing successive approximations to a target operant class (Cooper, Heron, & Heward, 1987). The operant classes targeted for change using shaping procedures have included compliance with medical treatment (Hagopian & Thompson, 1999), technical skill and performance in sports (Scott, Scott, & Goldwater, 1997), and communication (Lerman, Kelley, Vorndran, Kuhn, & LaRue, 2002), to name only a few. Despite the potential usefulness of shaping across a variety of response forms, research on procedural nuances and variations of shaping techniques is relatively stagnant. This is unfortunate, because several authors have noted that shaping techniques used in applied research

more often resemble an “art form” than an established procedure (Galbicka, 1994; Lattal & Neef, 1996; Platt, 1973). The lack of precision observed in implementation of shaping techniques is avoidable given the development of quantitative methods of shaping such as the percentile schedule of reinforcement (Platt, 1973).

The logic behind the percentile schedule is based on the general rules of shaping (Galbicka, 1994). Specifically, behavior must occur prior to being reinforced, so it is important to start the shaping procedure at a criterion for reinforcement within a range of behavior currently in an individual’s repertoire. In addition, behavior must be differentially reinforced toward a predetermined terminal criterion, such that there is a mixture of both extinction of and reinforcement of responding until the terminal criterion is reached. The percentile schedule follows these general rules and allows the specification of precise criteria for reinforcement throughout the shaping process. These criteria are based on the output of a mathematical equation:  $k = (m + 1)(1 - w)$ . In this equation,  $w$  denotes the density of reinforcement, and  $m$  is a fixed number of recent observations. The  $k$  parameter specifies what

---

We thank Cara Phillips for her assistance with collecting data for this project and Brian Iwata and Terry Scott for their comments on an earlier draft of this manuscript.

Address correspondence to Timothy Vollmer, Psychology Department, University of Florida, Gainesville, Florida 32611 (e-mail: vollmera@ufl.edu).

doi: 10.1901/jaba.2007.40-475

response value out of  $m$  most recently observed response values an upcoming response must exceed to satisfy the criterion for reinforcement. For example, with a  $w$  value of .5, a response meeting the criterion for reinforcement for a given observation will be observed half the time, offering a mixture of reinforcement and extinction of responses. An  $m$  value of 10 means that the 10 most recent observations are ranked according to their ordinal value from least to most, keeping the shaping procedure in touch with the individual's current repertoire. Solving the equation using these values gives a  $k$  value of 5.5. In most cases, it is easier to implement the percentile schedule by rounding the  $k$  value to a whole number. In the example, rounding  $k$  to 5 keeps the reinforcement density at approximately half of all responses and makes the criterion for reinforcement slightly less stringent than a larger  $k$  value. The  $k$  value of 5 denotes the observation ranked fifth among the 10 most recently observed and ranked observations is the value the current observation must exceed to meet the criterion for reinforcement.

A percentile schedule is sensitive to current levels of responding in that it allows continuous calculation of the criterion for reinforcement, using only the most recent observations (Galbicka, 1994). In addition, parameter values for the percentile equation can remain constant across clients; this keeps constant the overall probability that a response will be reinforced while allowing the reinforcement criteria to remain sensitive to idiosyncrasies in individual behavior. These elements of a percentile schedule help to enhance its sensitivity and precision and aid in the objective application of shaping techniques. Once parameter values for the percentile schedule have been selected, no further calculations must be made. Thereafter, the only job of a clinician is to rank recent observations and select the current reinforcement criterion. Given this, training clinicians to implement a percentile schedule does not require any teaching of the calculation of the

percentile schedule or even a conceptual understanding of the method. With or without such understanding, application of a percentile schedule allows an increase in the precision and consistency in application of shaping across clinicians and clients. An objective and consistent shaping technique such as this may be of prime importance in clinical settings. For example, it could be useful in cases in which shaping is being implemented with 1 client by several clinicians throughout the day and there is a concern that differences in technique could negatively affect acquisition of a skill. An objective and consistent shaping technique could also be of importance to researchers who would like tighter control over their participants' histories of reinforcement. The uniform application of shaping across conditions could also be of importance if one were doing a comparative analysis across two or more conditions using a multielement design. The potential usefulness of the method, therefore, indicates a need for research on its application and efficacy in clinical settings.

In research with nonhumans, percentile schedules have been used to examine inter-response times (Kutch & Platt, 1976; Platt, 1979), the effects of *d*-amphetamine on control of response number (Galbicka, Fowler, & Ritch, 1991), response acquisition (Galbicka, Kautz, & Jagers, 1993), and variable response sequences (Machado, 1989). Applications of the percentile schedule with humans have evaluated its efficacy in decreasing cigarette smoking (Lamb, Kirby, Morral, Galbicka, & Iguchi, 2004; Lamb, Morral, Galbicka, Kirby, & Iguchi, 2005; Lamb, Morral, Kirby, Iguchi, & Galbicka, 2004) and increasing variability in computer game playing (Miller & Neuringer, 2000). In the studies on decreasing cigarette smoking, a single observation was collected on 1 day, and reinforcement escalated across the study independent of the percentile schedule requirements. In addition, in the Lamb, Kirby, Morral, Galbicka, and Iguchi (2004) study,

participants were instructed as to what behavior was required to obtain reinforcement as part of a contingency-management program. Thus, rule governance may have been a major factor in the treatment effects. Further, the majority of studies have used adult participants. In the only investigation to date with children, Miller and Neuringer targeted reinforcing variability in computer game playing in adolescents with autism who engaged in stereotypy and fixed patterns of responding. The question remains, however, as to how effective percentile schedules might be at shaping a steady increase in academic behavior when instructions detailing the contingencies of reinforcement are not delivered and reinforcement of responding is determined solely by the percentile schedule.

Before implementing a percentile schedule of reinforcement, it is important to be aware of two formal assumptions involved in percentile schedules. First, behavior must be measured in a way that it can be assigned ordinal values and ranked according to those values. Second, those ranked values must not be sequentially related (Galbicka, 1994). The first assumption can be easily met by assigning numeric values to behavior. An example would be to rank the rate or duration of the behavior. The second assumption, however, requires that successive observations represent random samples that are independent of sequential dependencies. Sequential dependencies are cases in which the most recent response is dependent on a prior response (Galbicka). A commonly observed example of a sequential dependency is seen when a distribution of responses is bimodal or cyclical rather than independent and random. In such an example of a sequential dependency, the data can appear variable upon visual inspection. In light of the percentile schedule, a cyclical pattern in responding could make it impossible for the shaping process to advance. Suppose, for example,  $k = 3$  and the target behavior involves the number of math problems competed in 2-min trials. If in eight consecutive

trials an individual showed evidence of a cyclical pattern of responding, such as 1, 2, 3, 4, 1, 2, 3, 4 problems completed each session, the criterion for reinforcement would never advance under the percentile schedule.

The presence of sequential dependency may undermine the use of percentile schedules. It has been suggested, however, that the effects of sequential dependencies can be diminished by increasing the size of the comparison distribution ( $m$ ) (Galbicka, 1994; Platt, 1973). Galbicka presented a hypothetical situation similar to the one above and varied the size of the comparison distribution from 1 to 3 to 4. Using hypothetical data, Galbicka showed that a larger comparison distribution could decrease the effect of sequential dependencies and allow more effective shaping of behavior. However, a study that targeted smoking cessation indicated that at times a relatively smaller comparison distribution might be more effective in shaping behavior (Lamb et al., 2005). In a comparison of  $m = 4$  and  $m = 9$ , individuals exposed to a percentile schedule with  $m = 4$  reduced smoking more quickly than those exposed to a percentile schedule with  $m = 9$ . The researchers attributed this effect in part to an increased sensitivity to current levels of responding. It is also possible that the data were not sequentially related. More research is needed to determine the impact of sequentially related data on the efficacy of the percentile schedule.

There were two main purposes to the current study. The overall purpose was to examine the efficacy of a percentile schedule with students of varying skill levels, targeting increased durations of academic task engagement. A second purpose was to examine the  $m$  parameter of the percentile schedule to investigate the effects of various comparison distribution sizes on the efficacy of the percentile schedule as a method of shaping.

## METHOD

### *Participants and Setting*

Participants were 4 children who had been referred to our research program for interven-

tion due to low levels of compliance, defined as academic task engagement. Teacher reports indicated that the students did not respond to general verbal prompts to work, and that current classroom incentives were not working to increase the occurrence of independent task engagement.

Tony was a 9-year-old boy about to enter the fourth grade who had been diagnosed as learning disabled. Ashley was a 6-year-old girl in a multiaged exceptional student education classroom who had been diagnosed with other health disabilities and speech and language disabilities. Charles was a 14-year-old boy in a multiaged classroom who had been categorized as educable mentally handicapped. Anthony was a 9-year-old boy in second grade who had been diagnosed with specific learning disabilities including delays in speech, language, and fine motor skills.

Sessions were conducted at the participants' schools in a room resembling their classrooms. The room was equipped with two desks, a table, a whiteboard, three to five chairs, and materials used in the course of the study. Two to three sessions were conducted daily, 4 to 5 days per week. Each student worked alone, with an adult therapist and observers seated nearby. All observers remained out of the student's direct line of sight. Materials consisted of paper and pencils, preferred edible items, tokens in the form of poker chips, and a plastic bowl in which tokens were placed.

#### *Dependent Variables and Interobserver Agreement*

Tasks were selected based on teacher reports of what the students should be performing in the classroom. Tony's primary task was independent writing of sentences on a blank piece of paper in response to a journal topic (e.g., "What did you do this summer?"). Ashley's primary task was tracing letters of the alphabet outlined on a sheet of paper. Task engagement for Charles and Anthony was copying sentences on a lined sheet of paper. Erasing previously completed work counted as task engagement

because it allowed students to correct mistakes without observers counting such behavior as off task. There was an onset–offset criterion of 3 s for scoring occurrences of task engagement. Observers did not begin recording task engagement until 3 s of task engagement had passed, and did not cease recording its occurrence until 3 s with no task engagement had passed. This gave the participant time to manipulate the pencil to erase, to switch pages when one page of copying was completed within a session, or to pause briefly, without data collectors terminating an observation of on-task behavior prematurely.

Observers were graduate and undergraduate students who had previously attained three consecutive interobserver agreement scores of at least 90% with trained observers. Observers collected data during all sessions, some of which were videotaped for later scoring by additional observers. During sessions, observers were seated approximately 1.5 m away from the participant and therapist. Observers used hand-held computers to record real-time data of task engagement. Data were also collected on the therapists' delivery of tokens and prompts to determine procedural consistency.

To calculate interobserver agreement for compliance, data from each observer were divided into 10-s bins. For each bin, the smaller number of observed seconds of engagement was divided by the larger number and multiplied by 100% (Bostow & Bailey, 1969). The results were then averaged across the entire session. Interobserver agreement was assessed on 29%, 38%, 28%, and 25% of sessions for Tony, Ashley, Charles, and Anthony, respectively. Agreement on duration of on-task behavior averaged 95% (range, 91% to 100%) for Tony, 95% (range, 85% to 100%) for Ashley, 94% (range, 84% to 100%) for Charles, and 97% (range, 84% to 100%) for Anthony.

Treatment integrity was assessed on 29%, 38%, 28%, and 25% of sessions for Tony, Ashley, Charles, and Anthony, respectively.

Treatment integrity for token delivery was calculated by dividing the total number of tokens delivered after responses that met the reinforcement criteria by the total number of tokens delivered. Treatment integrity for prompt delivery was calculated by dividing the total number of correct prompts (prompts following 15 s of no task engagement) by the total number of prompts (Gresham, 1989). Treatment integrity for token delivery was above 96% (range, 97% to 100%), and prompt delivery was above 93% (range, 93% to 100%) for all participants.

#### *Preference Assessment*

Preference assessments were conducted for each participant using multiple-stimulus preference assessments without replacement (DeLeon & Iwata, 1996). Each assessment lasted 2 to 5 min and included eight edible items (candy). The six most preferred edible items were made available for token exchange in the experimental phases of the study.

#### *Pre-session Training and Baseline*

Prior to baseline sessions, brief tutorials were given to participants to demonstrate the required target behavior and the method of exchanging tokens. For example, when the task was to copy words from a prewritten sentence, a therapist modeled the appropriate behavior and then asked the participant to imitate that behavior. Appropriate responses were praised. Because the tasks were selected based on work the students were doing in the classroom, this was the extent of target behavior training. After a brief break, participants were told that tokens could be used to buy preferred candy. A therapist used a token to model the token exchange process. Participants were then given a noncontingent token and asked to exchange the token on their own. Each participant successfully completed the task and token tutorials in a single presentation.

Baseline sessions lasted 5 min for Tony, Ashley, and Anthony and 10 min for Charles.

Session durations were selected based on teacher preferences. Each participant was exposed to a minimum of three baseline sessions. At the beginning of each baseline session, the participant was presented with task materials and a blank piece of paper. The blank paper served as a potential distracter, and was made available in an attempt to mimic the classroom environment in which other activities might be available during any given academic task. Immediately prior to the start of a baseline session, the participant was told, "You can work if you want to. At the end of the session you can exchange any tokens you receive for candy." These instructions were delivered at the start of each session throughout the experiment. During baseline sessions, tokens were delivered noncontingently on a fixed-time (FT) 2.5-min reinforcement schedule. Under this schedule, two tokens were delivered in each 5-min session, and five tokens were delivered in each 10-min session. Token delivery consisted of the therapist placing a token into a bowl. The bowl was placed on the desk near the participant. In addition to tokens, verbal prompts to work (e.g., "time to work") were delivered on an FT 15-s schedule, contingent on the absence of task engagement. If task engagement occurred before 15 s elapsed, the timer was stopped and reset. At the end of sessions, smaller preferred edible items (e.g., Skittles®) were exchangeable for one token; larger edible items (e.g., Reese's Cups®) were exchangeable for two tokens.

#### *Percentile Schedule*

Each participant was exposed to a percentile schedule of reinforcement as described by Galbicka (1994), and 3 participants were exposed to parametric assessments of the percentile schedule. The effects of varying  $m$  values were studied using a reversal design whereby conditions in which the percentile schedule with a specific  $m$  value was in effect were presented alternately with baseline conditions. During each parametric analysis the value

Table 1  
Ranking of Recently Recorded Observations

Successive response durations (seconds)	Ranked response durations (seconds) ( $m = 5, k = 3$ )
12, 20, 10, 34, 14	10, 12, 14, 20, 34
(12) 20, 10, 34, 14, 19	10, 14, 19, 20, 34
(12), (20), 10, 34, 14, 19, 21	10, 14, 19, 21, 34

of  $w$  for each percentile schedule examined was set at .5. This value has been used in the successful implementation of percentile schedules in previous investigations (Galbicka *et al.*, 1991; Lamb, Kirby, Morral, Galbicka, & Iguchi, 2004). A value of .5 meant that half of a participant’s responses should meet the criterion for reinforcement. Each participant, with the exception of Anthony, was exposed to three different values of  $m$ . Due to time constraints, Anthony was exposed to a percentile schedule with only the largest value of  $m$  in effect. For all other participants, we examined  $m$  values of 20, 10, and 5 across conditions. We selected these values based on previous investigations of the percentile schedule (Galbicka *et al.*, 1991, 1993). The order of presentation of conditions across participants was semirandom. It was predetermined that the order of conditions would vary across participants, as a partial control for possible carryover and history effects. In addition, it was predetermined that each participant would finish his or her participation with exposure to the percentile schedule that showed the maximum treatment effect.

As an example of how the percentile schedule was implemented, given a comparison distribution of five, if the five most recent response

durations were 12 s, 20 s, 10 s, 34 s, and 14 s, these were ranked from least to most, yielding 10 s, 12 s, 14 s, 20 s, and 34 s. In this example, 10 s has a rank of 1, and 34 s has a rank of 5. When a new response was recorded, it was added to the ranked array and the oldest scored duration of task engagement in the array was removed from the ranking. For the previous example, the 12-s response value was discarded from the ranks when a new response duration was recorded (see Table 1 for a general example and Table 2 for an actual example with data from Tony’s first experimental session).

We used the most recent baseline observations to establish the initial criterion for reinforcement. If there were fewer than the required number of responses in baseline or if no responses occurred, we ranked however many responses had been observed and set the initial criterion for reinforcement at 3 s of task engagement until  $m$  observations of the behavior occurred. A value of 3 s was selected because it was the lowest possible given our method of scoring task engagement (with a 3-s onset and offset criterion). Once the required number of previous observations had been collected, the criterion for reinforcement was selected based on the rank specified by the percentile equation. For example, given the formula  $k = (m + 1)(1 - w)$  with  $w = .5$  and  $m = 20$ ,  $k$  is solved to equal 10. This meant that the value of the observation ranked 10th must be exceeded to meet the criterion for reinforcement. Given  $m = 10$ ,  $k = 5$ . Given  $m = 5$ ,  $k = 3$ . Each of the calculated  $k$  values was rounded to a whole number for ease of implementation. When an observed response exceeded the current criterion

Table 2  
Ranking of Recently Recorded Observations for Tony

Session	Successive response durations (seconds)	Ranked response durations (seconds) ( $m = 5, k = 3$ )
Baseline 1, 2, and 3	0, 43, 62, 45, 0	0, 0, 43, 45, 62
Session 1 Responses 1 through 6	(9), 4, 4, 3, 13, 4	3, 4, 4, 4, 13
Session 1 Responses 1 through 7	(9), (4), (4), 3, 13, 4, 12	3, 4, 4, 12, 13
Session 1 Responses 1 through 8	(9), (4), (4), 3, 13, 4, 12, 4	3, 4, 4, 12, 13

for reinforcement, a token was delivered in the same manner as in baseline. Verbal prompts to work were delivered as in baseline.

To simplify the application of the percentile schedule, a software program was developed that would run on the handheld computers used for data collection. This program ranked previously scored responses and identified the criterion for reinforcement. Specifically, the program allowed the therapist to input (a) the number of observations to rank and (b) the rank assigned as the criterion for reinforcement. For example, when  $m = 20$ , the therapist input 20 as the number of responses to be ranked and selected the 10th ranked response as the duration a future response must exceed to meet the criterion for reinforcement. The primary therapist indicated the start of the session with a verbal prompt of "1 2 3 start." At "start," the observer started the session on the handheld computer. When a participant displayed 3 consecutive seconds of task engagement, the observer started a timer displayed in the middle of the computer screen that visibly counted up from zero. When the duration of a response exceeded the reinforcement criterion, the timer value was highlighted red. The primary observer cued the therapist unobtrusively (e.g., with a slight nod of the head) to reinforce the response at its cessation. Occasionally, the therapist collected data herself, in which case such cues were unnecessary. The timer remained highlighted until the data collector stopped the timer (3 s after observing a cessation in task engagement), at which point the observation was automatically entered into the data file. The program ranked only the  $m$  most recent observations from the data file. This process was repeated with each response scored, until the end of a session. Observations were counted across sessions, such that the criterion for reinforcement at the start of a new session was based on  $m$  responses from the immediately preceding session.

A terminal criterion for the shaping procedure was selected based on teacher requests

concerning how long the participants were required to work independently in the classroom. For each participant, the terminal criterion was at least 80% of a session spent engaged in the task for a minimum of three consecutive sessions. A minimum total duration of task engagement of 4 min (240 s) per 5-min session was required for Tony, Ashley, and Anthony. The minimum for Charles was 8 min (480 s) per 10-min session.

## RESULTS

Figure 1 displays the results for Tony, Ashley, and Charles. Tony engaged in low levels of task engagement during baseline. Following baseline, he was exposed to a percentile schedule with  $m = 5$ . Under this schedule, there was an initial increase in his task engagement, but it varied across sessions. After reaching a relatively long duration of time spent engaging in the task in one session, responding decreased in subsequent sessions, and the predetermined terminal criterion was not reached. Immediately after introduction of baseline, all task engagement ceased. Close inspection of data from three immediately preceding sessions showed that Tony was not engaging in the task for approximately the first half of the percentile sessions. During the return to baseline, this lack of responding at the beginning of the sessions continued, allowing the delivery of a response-independent reinforcer, which may have suppressed responding. Following the reversal, a percentile schedule with  $m = 20$  was implemented. In this phase, responding increased rapidly, and the predetermined terminal criterion was met. Following this percentile schedule was another reversal to baseline, during which responding was maintained for several sessions before finally stabilizing with zero instances of task engagement. A percentile schedule with  $m = 10$  was then implemented and resulted in an initial increase in task engagement; however, responding was quite variable, and there appeared to be

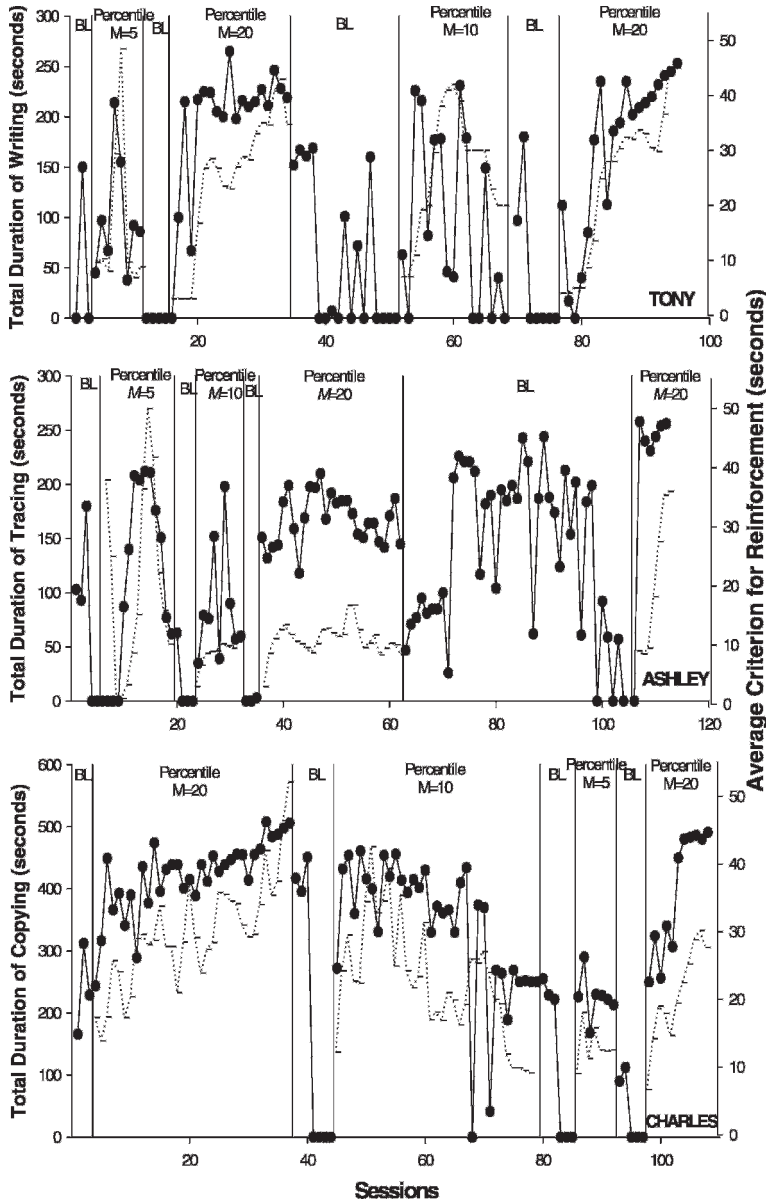


Figure 1. Results for Tony, Ashley, and Charles. Filled circles indicate the total duration of on-task behavior, plotted in seconds along the left y axis. Horizontal marks connected by a dotted line represent the average criterion for reinforcement of a response, plotted in seconds along the right y axis.

a downward trend in the data across sessions. In a final reversal to baseline, there was a relatively rapid decrease in responding. Tony's participation concluded with a final exposure to a percentile schedule with  $m = 20$ , during which the terminal criterion of 80% of the

session spent on task was met. The average criterion for reinforcement across each of the experimental phases showed increases and decreases that corresponded to increases and decreases in Tony's responding. As responding increased the criterion for reinforcement in-



creased. When responding decreased in subsequent sessions, there was a corresponding decrease in the criterion for reinforcement.

During Ashley's initial exposure to baseline conditions (Figure 1), she engaged in low levels of responding that decreased to zero. When a percentile schedule with  $m = 5$  was implemented, no increase in responding was observed initially. After several sessions, her responding increased; however, it soon decreased and the terminal criterion was not reached. During a reversal to baseline, responding decreased immediately to zero. Careful inspection of the data indicated that in the immediately preceding session during the percentile schedule, she quit working for approximately the last half of the session. After implementation of baseline, Ashley continued not to respond, allowing the delivery of a response-independent reinforcer. Following the reversal to baseline, a percentile schedule with  $m = 10$  was implemented. In this phase there was a more rapid initial increase in responding; however, responding was variable and the terminal criterion was not met. Following this phase was a reversal to baseline, during which there were immediate decreases in responding. A percentile schedule with  $m = 20$  was then implemented and was associated with a rapid increase in task engagement. Although the terminal criterion was not reached in this phase, response duration remained long and stable throughout the entire phase. In the final reversal to baseline, responding continued to occur at relatively long and variable durations across several sessions, but eventually decreased. Ashley's participation was concluded with a final exposure to a percentile schedule with  $m = 20$ , during which there was a rapid increase in task engagement, and she successfully reached the terminal criterion. The criterion for reinforcement plotted in each of the experimental phases for Ashley showed increases and decreases that corresponded to increases and decreases in Ashley's responding, similar to what was observed with Tony.

During Charles' initial exposure to baseline (Figure 1) task engagement occurred at or below half the duration of each session. His percentile schedule began with  $m = 20$  and was associated with a steady increase in duration of task engagement across sessions. There was a slight disruption in responding after Charles had a week-long school break between Sessions 10 and 11. Thereafter, his responding continued to increase, and the terminal criterion for task engagement was met. After reversing to baseline, a percentile schedule with  $m = 10$  was implemented, during which durations of task engagement were initially long but showed an overall downward trend; responding eventually stabilized at durations of less than half of each session spent engaging in the task. Following a reversal to baseline, a percentile schedule with  $m = 5$  was implemented, during which task engagement stabilized with less than half the duration of a session being spent engaging in the task. Charles' final condition, a percentile schedule with  $m = 20$ , was associated with a rapid increase in responding, and the terminal criterion was reached. The criterion for reinforcement plotted in each of the experimental phases for Charles showed increases and decreases that corresponded to increases and decreases in responding, similar to what was observed with Tony and Ashley.

Data for Anthony are depicted in Figure 2. Anthony engaged in low to zero durations of task engagement during the initial baseline. After implementation of the percentile schedule with  $m = 20$ , there were initially variable durations of responding across sessions; however, durations of task engagement eventually increased. There was an anomalous decrease in duration of on-task behavior at the 48th and 49th sessions when, at the end of sessions, Anthony informed us he was ill. He subsequently missed 2 days of school. After his return, his responding returned to previously long durations of task engagement, and he ultimately met the terminal criterion. During

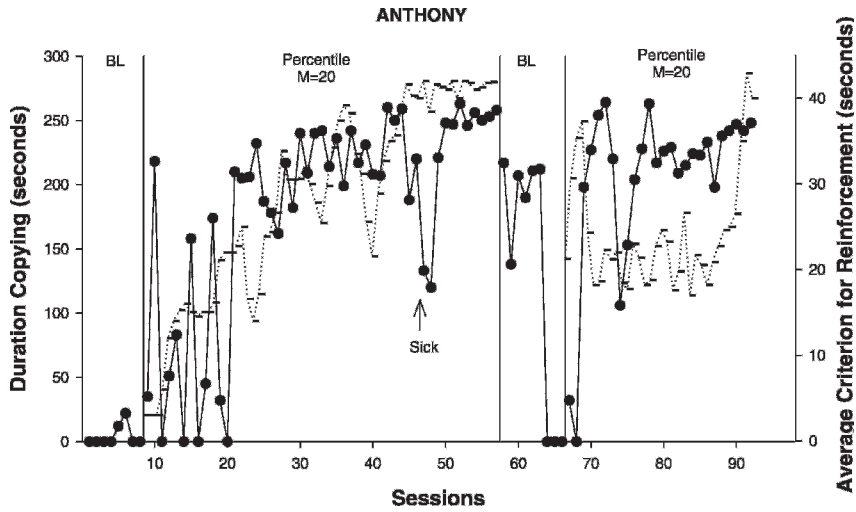


Figure 2. Results for Anthony. Filled circles indicate the total duration of on-task behavior, plotted in seconds along the left y axis. The horizontal marks connected by a dotted line represent the average criterion for reinforcement of a response, plotted in seconds along the right y axis.

a reversal to baseline following the percentile phase, response durations initially remained relatively long before eventually decreasing to zero. Following the baseline reversal, there was a final replication of the percentile phase ( $m = 20$ ). Anthony's responding was variable initially, although the terminal criterion was reached in a shorter time than was required during his first exposure to a percentile schedule. The average criteria for reinforcement plotted in each of the experimental phases showed increases and decreases that corresponded to increases and decreases in responding. These results were similar to those observed with Ashley, Tony, and Charles.

## DISCUSSION

We examined the effectiveness of a percentile schedule to increase task engagement, using three values of  $m$ . Results indicated that the percentile schedule was effective when a relatively large number of previous observations was taken into account. The current examination is the first to date using percentile schedules to shape academic behavior. In addition to extending the generality of the percentile method

of shaping, the use of a token economy offered an extension of the validity of previous research. Conditioned reinforcers have been used during previous examinations of percentile schedules, in which adults were given money following smoking omission periods (Lamb, Kirby, Morral, Galbicka, & Iguchi, 2004; Lamb, Morral, Kirby, Iguchi, & Galbicka, 2004). Given our diverse population of participants, edible items were the most universal reinforcers to make contingent on meeting the criteria. Unfortunately, delivery of edible items during a session could interrupt on-task behavior because of time spent in consumption. Tokens permitted consumption to be postponed.

The results of the current experiment were similar to results presented by Lamb, Morral, Kirby, Iguchi, and Galbicka (2004), in which certain manipulations of  $w$  were more effective at shaping decreases in smoking in adults. Their findings suggest that parametric values can have a substantial impact on the efficacy of the percentile schedule as a method of shaping. Lamb *et al.* (2005) also presented results related to the current experiment when they showed that individuals exposed to a percentile schedule where  $m = 4$  reduced their smoking more quickly than those exposed to a percentile

schedule where  $m = 9$ . The finding seems to contradict results of our study, but actually both studies point to a role of the  $m$  value. That Lamb et al. (2005) indicated a relatively small value for  $m$  to be most effective might stem from differences in methodology. For example, the current investigation examined shaping increases in academic behavior, whereas the Lamb et al. (2005) investigation examined the percentile schedule as a method of decreasing cigarette smoking. It is also possible that when shaping increases behavior, a percentile schedule with a larger  $m$  value is more effective, and that when shaping decreases in behavior, a percentile schedule with a smaller  $m$  value is more effective. Alternatively, it is possible that sequential relatedness in the data may be more of an issue when collecting large amounts of data in a day. Autocorrelation is a mathematical tool that is useful for finding repeating patterns in a signal. We used an autocorrelation analysis to test for the presence of sequential dependencies in a random sample of data from each participant across experimental phases. Although the data are too lengthy to present in the current study, we found the data in each sample to be sequentially related (data are available from the first author). The analysis showed that increasing the window of observations, however, decreased the data correlation, resulting in fewer cyclical patterns of responding that may slow down the shaping procedure.

An alternative explanation of the data could be that when the relatively small  $m$  value of 5 was in effect, sudden increases in the duration of responding could alter the criterion for reinforcement within five responses. In such cases, if all responses did not suddenly fall within the larger durations, the criterion for reinforcement would not be met. When the larger  $m$  value of 20 was in effect, however, rapid changes in responding did not alter the criterion for reinforcement for 20 responses. In this phase, if all responses did not suddenly fall in the larger durations, the criterion for

reinforcement was still likely to be met. An example of this can be seen in Tony's data. In Session 7, with the  $m = 5$  percentile schedule, there was a substantial increase in the total duration of writing. With this increase there was a substantial and rapid increase in the average criterion for reinforcement. In this session Tony met the criterion for reinforcement on only three of seven opportunities. In Session 18, with the  $m = 20$  percentile schedule, there was an identical increase in the total duration of writing. Unlike the session with a lower  $m$  value, however, there was not an immediate increase in the average criterion for reinforcement. In this session, Tony met the criterion for reinforcement on six of the seven opportunities. There was an overall greater delivery of reinforcement in the percentile schedule with  $m = 20$ , which may have influenced the efficacy of the schedule.

Given the multiple differences between this investigation and previous investigations of the percentile schedule, further research is required to empirically determine which methodological differences resulted in the observed differences across investigations. Overall, however, the results of Lamb et al. (2005) in relation to the current findings suggest that variations in parameter values may alter the efficacy of the percentile schedule of reinforcement in different ways across various procedures. We can tentatively recommend, based on the Lamb et al. (2005) findings and the current results, that when few observations are collected in a day, a relatively small window of observations may be taken into account to keep reinforcement criteria sensitive to current changes in responding. When numerous observations are collected in a day with that participant, however, a larger window of observations may be more effective at shaping the target behavior.

A novel result observed with each of the participants in this experiment was the initial insensitivity to the noncontingent delivery of tokens in baseline following exposure to

a percentile schedule with  $m = 20$ . For each participant, responses were at their longest stable durations during conditions in which  $m = 20$ . In the unsignaled transition to baseline sessions, responding remained at longer durations relative to baseline conditions that followed percentile schedules in which there were lower, or more variable, levels of responding. This may be in part due to the fact that although tokens were delivered noncontingently, token delivery often occurred during task engagement. The FT schedule of reinforcement throughout baseline sessions was relatively thin, however, and responding eventually decreased. It is possible that if noncontingent reinforcers were delivered on a denser schedule, the contiguous pairing of token delivery with instances of task engagement could have increased and responding could have been maintained. This has implications for ease of teacher application of the intervention. It is possible that, following successful completion of a percentile method of shaping, the teacher could deliver reinforcers on an FT schedule that resembles the response-dependent schedule and maintain the target behavior. This effect warrants further investigation.

Another effect observed in this experiment was that increases in task engagement ultimately resulted in decreases in the delivery of tokens. This was a product of our method of scoring durations of task engagement. Because a token was delivered only at the end of an instance of task completion, longer instances decreased the opportunity to receive tokens. In conditions in which  $m = 20$ , the changes in behavior were less variable than when  $m = 5$  or  $m = 10$ . This lack of response variability resulted in a more gradual thinning of token delivery. The gradual thinning of reinforcement in this phase may have contributed to the initial insensitivity to the FT schedule present during the baseline that followed it. In addition, the gradual thinning of reinforcement was potentially beneficial, because participants were not left on unreasonably

dense schedules of reinforcement prior to completion of the study. This made it more feasible to place the students back in their classrooms in which there was a less dense schedule of reinforcement in effect.

A limitation to the current study involved the method of scoring task engagement. To best capture instances of working, there was a 3-s onset–offset criterion. In sessions in which a percentile schedule was in effect, this resulted in a 3-s delay to token delivery when task engagement ended. On the few occasions that task engagement was not quickly followed by additional task engagement, a token was delivered in the absence of task engagement. The delivery of a token in the absence of responding could have resulted in adventitious reinforcement of the absence of responding. Overall, however, this element of our procedure did not appear to preclude the shaping of task engagement to some degree in all experimental phases and to the terminal criterion when a percentile schedule with  $m = 20$  was in effect.

The potential utility of a percentile schedule for clinicians and applied researchers remains to be seen. However, the procedure appears to be promising in many respects. For example, similar to the use of the percentile schedule to shape decreases in cigarette smoking (Lamb, Kirby, Morral, Galbicka, & Iguchi, 2004; Lamb, Morral, Kirby, Iguchi, & Galbicka, 2004), the percentile schedule could be used to shape decreases in problem behavior as well as thin the delivery of reinforcers in common procedures such as the differential reinforcement of other (DRO) or of low-rate behavior. When using a DRO procedure, the ordinal quantity targeted by the percentile schedule could be the duration of time a specific response does not occur. Procedurally, the thinning of reinforcement during DRO could be conducted in a manner similar to the one used in the current shaping of academic engagement. Specifically, the duration of intervals without problem behavior could be ranked and the

criterion for reinforcement determined based on these rankings. In this example, the criterion would indicate the amount of time a problem behavior would have to not occur for other behavior to be reinforced.

Another application of the percentile schedule could involve the teaching of self-care skills such as oral hygiene, clothes washing, or dish washing. Using oral hygiene as an example, a task analysis of tooth brushing similar to the one presented by Horner and Keilitz (1975) could be used to assign ordinal values to each step of the complex skill. If picking up the toothbrush was Step 1, placing toothpaste on it was Step 2, placing it in mouth was Step 3, and so on, these numbers could be ranked to determine what upcoming step must be completed to meet the criterion for reinforcement (cf. Galbicka, 1994). A benefit to the use of a percentile schedule for teaching self-care skills is that in cases in which the task analysis includes relatively few steps, or when each step requires a substantial input of time on the part of the student, then the ranking of responses could be done by the teacher or clinician without requiring the aid of technology.

Shaping is a powerful method available to practitioners who attempt to promote changes in current behavioral repertoires; however, shaping can be complicated if decisions must be made quickly. For example, which responses should be reinforced? How quickly should the criterion for reinforcement be increased? How large an increase in this criterion should be made? What should happen when the learner has a setback? (Galbicka, 1994). Some clinicians and researchers may be quite skilled at making these decisions, and shaping may progress rapidly. Other clinicians and researchers, however, may not be as skilled, and rarely will any two be exactly alike in their approach to the technique. The percentile schedule offers a method of shaping that can remove the need to make sudden within-session decisions and standardizes the shaping procedure across

therapists and clients. Once decisions on the values of  $m$  and  $w$  are made and the value of  $k$  is obtained, no additional mathematical computations are required. In addition, with the aid of a computer program such as the one used in this study to automatically rank responses and designate the criterion of reinforcement, response effort is decreased. It is possible to develop a simple Excel spreadsheet that will rank recent observations and highlight or otherwise allow one to easily identify the criterion for reinforcement. We did in fact develop such a program, and it is available from the first author. Although such a tool could help to ease application of the technique, the method is sufficiently complicated that if prompting and contingent reinforcement are adequate in establishing the response and consistency in technique across therapists is not of interest, a percentile schedule of reinforcement would not be recommended. Other control techniques, such as a changing criterion design (Hartmann & Hall, 1976), may be more suitable for use with shaping in such circumstances. The percentile method, however, offers sufficient benefits under certain conditions that its consideration as a therapeutic intervention and further research into the generality of its application are warranted.

## REFERENCES

- Bostow, D. E., & Bailey, J. B. (1969). Modification of severe disruptive and aggressive behavior using brief timeout and reinforcement procedures. *Journal of Applied Behavior Analysis, 2*, 31–37.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Englewood Cliffs, NJ: Prentice Hall.
- DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519–533.
- Galbicka, G. (1994). Shaping in the 21st century: Moving percentile schedules into applied settings. *Journal of Applied Behavior Analysis, 27*, 739–760.
- Galbicka, G., Fowler, K. P., & Ritch, Z. J. (1991). Control over response number by a targeted percentile schedule: Reinforcement loss and the acute effects of *d*-amphetamine. *Journal of the Experimental Analysis of Behavior, 56*, 205–215.

- Galbicka, G., Kautz, M. A., & Jagers, T. (1993). Response acquisition under targeted percentile schedules: A continuing quandary for molar models of operant behavior. *Journal of the Experimental Analysis of Behavior*, *60*, 171–184.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, *18*, 37–50.
- Hagopian, L. P., & Thompson, R. H. (1999). Reinforcement of compliance with respiratory treatment in a child with cystic fibrosis. *Journal of Applied Behavior Analysis*, *32*, 233–236.
- Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, *9*, 527–532.
- Horner, R. D., & Keilitz, I. (1975). Training mentally retarded adolescents to brush their teeth. *Journal of Applied Behavior Analysis*, *8*, 301–309.
- Kutch, D. O., & Platt, J. R. (1976). Reinforcement rate and interresponse time differentiation. *Journal of the Experimental Analysis of Behavior*, *26*, 471–486.
- Lamb, R. J., Kirby, K. C., Morral, A. R., Galbicka, G., & Iguchi, M. Y. (2004). Improving contingency management programs for addiction. *Addictive Behaviors*, *29*, 507–523.
- Lamb, R. J., Morral, A. R., Galbicka, G., Kirby, K. C., & Iguchi, M. Y. (2005). Shaping reduced smoking in smokers without cessation plans. *Experimental and Clinical Psychopharmacology*, *13*, 83–92.
- Lamb, R. J., Morral, A. R., Kirby, K. C., Iguchi, M. Y., & Galbicka, G. (2004). Shaping smoking cessation using percentile schedules. *Drug and Alcohol Dependence*, *76*, 247–259.
- Lattal, K. A., & Neef, N. A. (1996). Recent reinforcement-schedule research and applied behavior analysis. *Journal of Applied Behavior Analysis*, *29*, 213–230.
- Lerman, D. C., Kelley, M. E., Vorndran, C. M., Kuhn, S. A. C., & LaRue, R. H., Jr. (2002). Reinforcement magnitude and responding during treatment with differential reinforcement. *Journal of Applied Behavior Analysis*, *35*, 29–48.
- Machado, A. (1989). Operant conditioning of behavioral variability using a percentile reinforcement schedule. *Journal of the Experimental Analysis of Behavior*, *52*, 155–166.
- Miller, N., & Neuringer, A. (2000). Reinforcing variability in adolescents with autism. *Journal of Applied Behavior Analysis*, *33*, 151–165.
- Platt, J. R. (1973). Percentile reinforcement: Paradigms for experimental analysis of response shaping. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 7. Advances in theory and research* (pp. 271–296). New York: Academic Press.
- Platt, J. R. (1979). Interresponse-time shaping by variable-interval-like interresponse-time reinforcement contingencies. *Journal of the Experimental Analysis of Behavior*, *31*, 3–14.
- Scott, D., Scott, L. M., & Goldwater, B. (1997). A performance improvement program for an international-level track and field athlete. *Journal of Applied Behavior Analysis*, *30*, 573–575.

Received June 20, 2006

Final acceptance February 22, 2007

Action Editor, Louis Hagopian