A Technology for Evaluating Relations between Response Frequency and Academic Performance Outcomes*

Kimberly Nix Berens, Thomas E. Boyce, Nicholas M. Berens,
Janice K. Doney, and Amy L. Kenzer

University of Nevada- Reno

Since the 1970's, proponents of Precision Teaching have cited the relation between fluent performance and various long-term academic performance outcomes. Unfortunately, clinical work has mainly been used to demonstrate the importance of high correct response frequencies in education. The current paper describes a technology currently being used in a university-based Precision Teaching center to systematically evaluate relations between response frequency and academic performance outcomes with school children. Specifically, three studies are reported which evaluate retention, endurance, and application as a function of response rate during training. Results indicate that higher rates of responding during training yielded better performance during tests of retention, endurance, and application. The studies are discussed in terms of technological innovations for educational reform and directions for future research.

DESCRIPTORS: Precision Teaching, fluency, retention, endurance, application

In the Precision Teaching (PT) literature, Retention Endurance Application Performance Standards (REAPS) have been described as the positive outcomes to fluent performance. From this perspective, fluency is defined as a mastery criterion that requires both accuracy and speed (i.e., high correct response frequencies). Proponents of fluency-based instruction suggest that fluent performers are more likely to maintain their accurate, high-rate performances over periods of time without practice (retention) and over long durations in the presence of distractions (endurance/stability), as well as acquire complex skills more quickly and easily (application/adduction) (Binder, 1996; Bloom, 1986; Johnson & Layng, 1992; Lindsley, 1971).

As described above, REAPS could have great significance for learners in traditional education settings. First, increased skill retention following extended school absences (e.g., summer break) would require that less time be spent reviewing skill objectives from the previous school year. Second, behavioral endurance during standardized tests could enhance student performance by decreasing distractibility and fatigue. Finally, application outcomes could increase the likelihood that learners would continue to excel in and pursue advanced coursework during their secondary or post-secondary educations. Overall, it appears that if fluency leads to REAPS, fluency-building strategies may be used to complement more traditional methods of instruction as a means of producing better learning gains.

The notion of fluency extends beyond the PT literature. Proponents of curriculum-based measurement (CBM) also use behavior frequency as a basic unit of measurement for evaluating academic proficiency (Deno, 1985; Fuchs, Deno, & Mirkin, 1984; Marston, Mirkin, & Deno, 1984). CBM maintains similarities to PT in that count per time per time, or responses per minute per week, serves as the basic datum from which educational decisions are made. Additionally, from both perspectives, the higher the response frequency the more proficient the learner is considered to be. Research in CBM indicates frequency of correct response is as reliable a measure of academic proficiency as commercially available standardized test scores, and tends to provide educators with more information regarding the specific nature of skill deficits (Elliot & Fuchs, 1997; Fuchs & Deno,

*The authors would like to thank Adel C. Najdowski for her contributions to the initial data analyses, which eventually gave rise to this line of research. We would also like to graciously thank Rebecca Flores for her valuable assistance in data collection and implementation of the research protocols. Additional thanks are extended to all of the CAL Tutors and Case Managers, who help make the scientist-practitioner model of clinical practice a reality for our center.

*Correspondence concerning this article should be addressed to Kimberly Nix Berens, Department of Psychology/296, University of Nevada-Reno, Reno, NV 89557 or Thomas E. Boyce, Department of Psychology/296, University of Nevada-Reno, Reno, NV 89557. Electronic mail may be sent to Kimbo08@aol.com or teboyce@unr.edu

1992; Wesson, Skiba, Sevcik, King, & Deno, 1984; Wilson, Schendel, & Ulman, 1992). However, CBM defines fluency as frequencies of correct responses that fall at or above the class average (Hintze, Conte, Shapiro, & Basile, 1997). Therefore, less emphasis is placed on specific performance outcomes obtained with individual learners when compared with PT methods.

Although there are differences between how PT and CBM use fluency as a measure of learning, it is generally agreed upon that fluent performances result in long-term benefits to the learner. Unfortunately, much of the evidence for REAPS has been obtained clinically (Binder, 1996). Although clinical evidence is important and should not be overlooked, it appears that more rigorous empirical work evaluating the relation between increasing response frequencies and REAPS is sorely needed. In order to influence a shift in traditional educational practices in favor of behavior analytic approaches that emphasize frequent measurement and true skill mastery, outcome data across large numbers of learners is needed.

The current paper describes a technology currently being used to examine relations between response frequency and REAPS within a university-based PT center. The primary mission of the center is to design effective instructional strategies for improving the academic performances of students in grades K-12. The center is also committed to the use of a scientist-practitioner model of psychological practice where data are systematically collected for use in: (a) making educational decisions, (b) conducting general program evaluation, and (c) disseminating scientific information regarding relations between response frequency and academic performance outcomes.

GENERAL CENTER OVERVIEW

Students

Twenty-five students in grades one through 10 currently attend the Center for Advanced Learning (CAL) for 2 to 6 hours per week and receive tutoring in mathematics and reading. Each student's curriculum is tailored to meet his/her specific needs and is established from performance on a comprehensive basic skills assessment. Each skills assessment is curriculum-based, such that tests are generated from local grade-level curricula. However, these curricula are broken down into basic components so that only finite skills are being assessed. Learners' correct responses per minute are initially evaluated for grade-level curricula. However, if a learner's correct responding falls below the established fluency aims for any grade-level skills, their performances on skills from previous grade levels are evaluated until fluent performances are obtained. Skills at the next level are then targeted for fluency training.

Curricula

Similar to the basic skills assessment, the CAL math and reading curricula are composed of various programs that represent overall skill objectives for grades one through 12. These programs are broken down into the basic component skills required for performance of the overall skill objective, or composite skill. For example, the phonetic reading program is broken down into (a) identifying letters, (b) identifying phonics sounds, (c) identifying vowel blends, (d) identifying consonant blends, (e) identifying three-letter words, and (f) identifying words for grade levels 1-12 (all words can be identified through word-attack skills). The stimuli associated with a particularlevel in a program are presented on various worksheets or sets of flashcards. This levels system applies for all math and reading programs included in the curriculum, with various worksheets and flashcards associated with each level.

A session book is generated for each student. The book includes: (a) a curriculum checklist identifying the component skills and overall composite skill objectives to be mastered during the student's tutoring sessions, (b) paper copies of each skill level in a program to be conducted during tutoring sessions, (c) standard celeration charts (Pennypacker, Koenig, & Lindsley, 1972) for recording correct and incorrect response frequencies obtained during each program, and (d) a log for communicating program changes, directing questions to the Case Managers/Advisors, and describing problems or highlighting achievements that occur during a student's session. Additionally, during fluency training sessions, tutors are equipped with: (a) digital timers for the accurate monitoring of various timing periods, (b) handheld counters for tallying rapid response frequencies, (c) pencils, erasers, and rulers for charting, and (d) boxes of highly preferred stimuli for use as rewards.

GENERAL TRAINING PROCEDURES

Fluency Training

Skill introduction. Skill introduction is the first step in fluency training. Specifically, each student's correct response frequency on an academic task is evaluated during 15-s, 30-s, and 1-min timings. The timing length that produces response frequencies closest to the fluency range determines the interval used for initial training.

For example, if a student engages in higher frequencies on a task during a 15-s timing versus a 30-s or 1-min timing, then frequency building during 15-s timings is introduced. Thirty-second timings are introduced after fluent responding is achieved at 15-s and so on until the learner is performing fluently during 1-min timings.

Frequency building. Following skill introduction, frequency building during the previously selected timing length commences. During frequency building, each student is instructed to perform the skill as quickly as possible. Students are periodically prompted during the timing to respond more rapidly (e.g., "Go faster" or "Hurry"), if pausing or decrements in frequency are observed during the interval.

Tutors deliver verbal praise during the course of a timing when a student makes significant improvements in accuracy or frequency as compared to his/her previous timing (e.g., "Awesome" or "Great"). Errors are noted, but corrective feedback is withheld until the end of the timing. At that time, the tutor initiates correction trials for all incorrect items.

A correction trial consists of the tutor instructing the student to respond to the incorrect item again, providing immediate praise for a correct response or immediate corrective feedback for an additional incorrect response. Corrective feedback entails the tutor providing the student with the correct response (i.e., modeling), and then requiring that the student repeat the correct response. The tutor then requires that the student respond to the item again, and provides immediate praise for the correct response. Correction trials of this sort are repeated until the student can emit independent correct responses to all incorrect items.

Throughout the course of the training session, students are eligible to receive verbal, tangible, or edible rewards contingent upon various response requirements. The specific contingency implemented with each student depends upon nuances specific to that individual. For example, some students show rapid increases in response frequencies across days (i.e., high celerations). As a result, these students receive reinforcement for doublings in performance (i.e., X2.0 celeration) throughout the week. However, some students require that reinforcement be made contingent upon accurate responding before increases in frequency can be targeted. Other students require that reinforcement be made contingent upon on-task or compliant behavior.

In addition, it is standard practice in CAL to reinforce a student's performance if fluency is achieved during a given timing length and he/she proceeds to a longer timing length (i.e., increases in endurance). In order for a student to proceed to a longer timing length, his/her response frequency per minute must fall within the fluency range (i.e., divided by .25 and .5 for a 15-s and 30-s timing, respectively). Finally, performance is reinforced for Phase 1 and Phase 2 mastery of a skill.

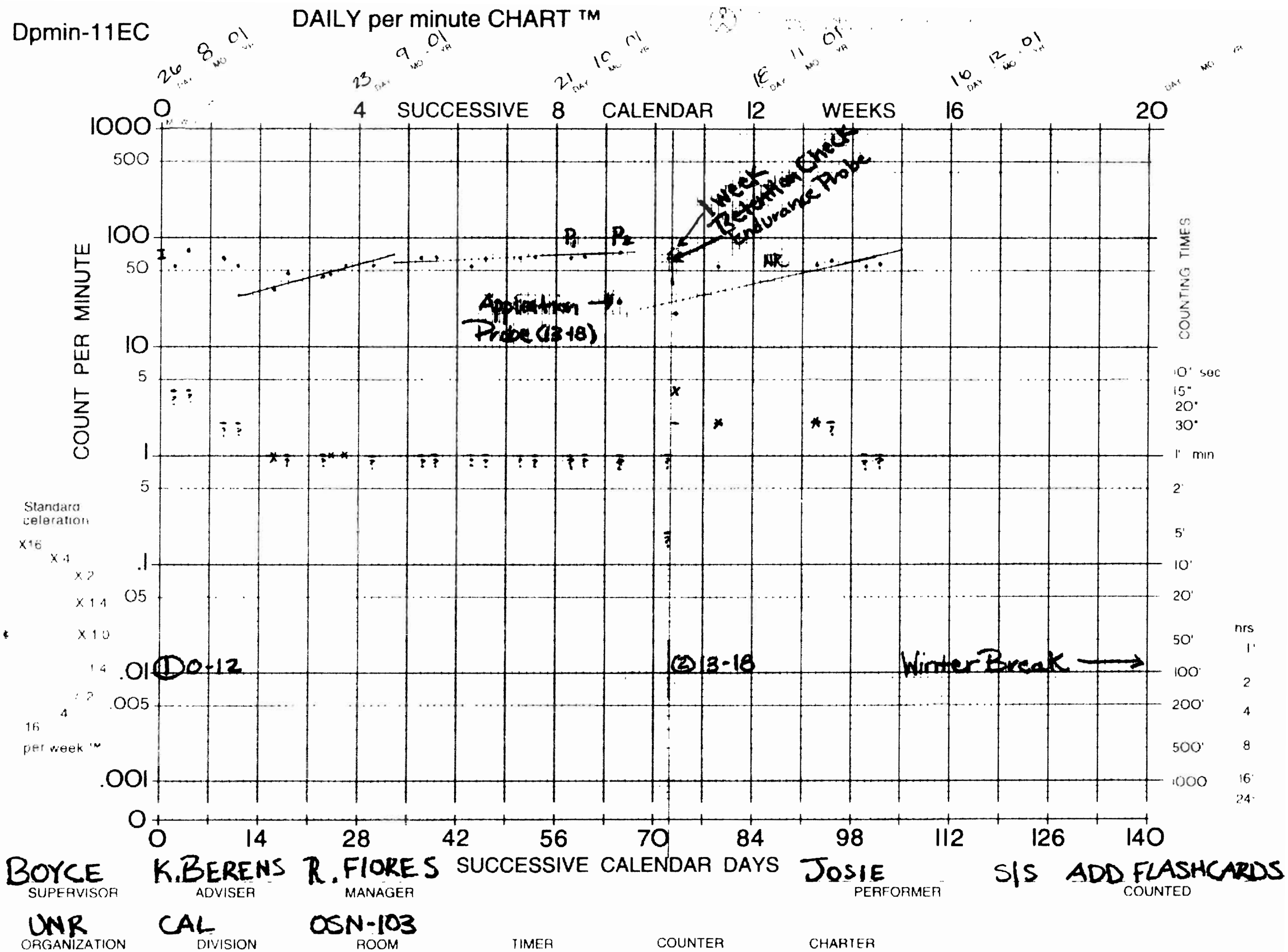
Phase 1 mastery is achieved when a correct response frequency falls within the fluency range for two consecutive 1-min timings during a session. Phase 2 mastery is achieved when a correct response frequency falls within the fluency range during the first timing conducted on that skill during the first session after Phase 1 mastery has been achieved (generally 2 days apart). If the student fails to achieve Phase 2 mastery, then he/she must restart the mastery sequence (i.e., achieve Phase 1 and Phase 2 mastery on consecutive sessions).

Since individualized programs of instruction are implemented with all students as a means of enhancing learning, the numbers of timings conducted on each skill during each session depend upon individual performance characteristics. Therefore, response opportunities vary across students as a function of their individual performance during sessions, and are evaluated as part of the data analysis in each of studies reported below.

General Data Collection Procedures and Calculation of IOA

For all students, frequency of correct responses, incorrect responses, and skipped items are recorded during sessions using standard celeration charts. A standard timings chart is used to record the response frequencies obtained during all timings conducted on a particular skill during a session. After all timings have been conducted on that skill, the highest correct response frequency obtained for that skill area is then plotted on a standard daily chart. Figure 1 shows an actual daily chart from one of our learners, who is also included in the analyses reported below.

Secondary observers obtain point-by-point agreement measures during a portion of all tutoring sessions. At the completion of the analyses reported below, secondary observers had obtained agreement measures for 782 timings conducted throughout the 2001-2002 school year and 2002-summer session. Exact agreement scores were calculated from these measures by first identifying the total number of agreements and disagreements obtained during the timing. Agreements included total number of correct items, incorrect items, and skipped items in which the two observers reported the same score. Disagreements included the num-


Dpmin-11EC
DAILY per minute CHART ™
SUCCESSIVE CALENDAR WEEKS
COUNT PER MINUTE
COUNTING TIMES
Standard celeration
per week ™
SUCCESSIVE CALENDAR DAYS
1 week Retention Check
Endurance Probe
Application Probe (13-18)
P1
P2
NR
① 0-12
② 13-18
Winter Break
BOYCE
SUPERVISOR
K. BERENS
ADVISER
R. FLORES
MANAGER
JOSIE
PERFORMER
S|S ADD FLASHCARDS
COUNTED
UNR
ORGANIZATION
CAL
DIVISION
OSN-103
ROOM
TIMER
COUNTER
CHARTER

ber of correct, incorrect, or skipped items for which conflicting scores were reported. Exact agreement coefficients were then calculated for each timing by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100%. An overall agreement score across all timings was then calculated. The average agreement score across the 782 timings was 96.8%, and ranged from a low of 78% to a high of 100%.

STUDY 1: RETENTION

Method

Participants

Five students enrolled in CAL were selected for inclusion in the retention analysis. These 5 participants were included due to their having curricula with at least two programs within a similar academic domain having the same fluency aim (65 or more responses/min), thus allowing for appropriate within and between subject comparisons.

Jack (7th grade), Emily (6th grade), Larry (7th grade), and Josie (10th grade) were typically developing students receiving instruction in general education classrooms at private schools in northern Nevada. Jack and Emily were enrolled in CAL to receive skill building in mathematics. Larry and Josie were enrolled to receive skill building in reading and mathematics.

Karen was a 4th grade female with mild Downs Syndrome in special education at a public school in northern Nevada. Karen was enrolled in CAL to receive skill building in reading, mathematics, and fine motor skills. She was also receiving intervention through CAL for disruptive classroom behavior (e.g., out-of-seat behavior, noncompliance, etc.).

Responses

For all participants, correct responses per minute on basic computation flashcards were included in the retention analysis. Basic computation included: (a) addition facts (0-18), (b) subtraction facts (0-18), (c) multiplication facts (0-12), (d) division facts (0-12), (e) reducing fractions, and (f) converting improper fractions to proper fractions. All basic computation flashcard tasks entailed the same fluency aim of 65 or more correct responses per minute. The fluency aim for this skill was established through normative sampling (Binder, 1996), where the number of computation flashcards completed per minute was evaluated across a series of timings conducted with exemplary learners, undergraduate tutors, and graduate students affiliated with our center. The distribution of scores obtained on this skill area across individuals was evaluated to produce a range of optimal frequencies, with 65 facts spoken per minute serving as the low-end of the range.

Reinforcers

All 5 participants received edible rewards (i.e., candy or snacks) for: (a) any increase in response frequency over the previous session, (b) achievement of fluent performance during shortened timing lengths (e.g., 15-s. or 30-s. timings), (c) achievement of Phase 1 mastery, and (d) achievement of Phase 2 mastery. Karen also received edible rewards for staying on-task and following instructions throughout her sessions (these were delivered randomly at the discretion of her tutor).

Retention Probes

Retention probes were conducted on mastered and unmastered skills following a naturally occurring 1-month break between semesters (i.e., winter break). During probes for retention, participants were given up to four opportunities to perform the skill as a means of controlling for warm-up effects. A probe session could involve fewer than four timings if the participant responded fluently for two consecutive timings. For those participants who had achieved Phase 2 mastery of a skill prior to the retention period, failure to perform fluently during two consecutive probe timings resulted in that skill being reintroduced during training.

Data Analysis

Data for the 5 participants were analyzed in terms of median training frequency and total number of responses emitted during training. The relation of these measures to the percentage of retention was explored.

Median training frequency. The median training frequency was calculated from all of the participants' training data for each separate skill. To calculate the median frequency for a skill, each score obtained on that skill prior to the retention period was entered into a computer database that automatically calculated the median measure for the distribution of values. This median served as a summary measure of the participant's training performance for use in comparing training performance with the percentage of retention.

Total responses emitted. A total responses emitted score was calculated from each participant's training data to serve as a measure of practice. This score was calculated for each participant by summing all of the response frequencies obtained across all timings prior to the retention period. The total responses emitted calculation reflected actual responses. Therefore, the numbers of responses obtained during shortened time periods (i.e., 15-s or 30-s) were not reported as count per minute for this calculation.

Percentage of retention. Percentage of retention was calculated for each participant's performance of a skill during retention probes. To obtain a more comprehensive measure of performance before and during the probe, the percentage of retention calculation involved dividing the average frequency obtained during the retention probe (i.e., summing each frequency obtained during the probe and dividing by the total number included in the sum) by the average of the last two training points obtained before the probe. This computation yielded a percentage of each participant's previous response frequency that was maintained following a period without instruction or practice.

RESULTS

The relation between each participant's median training frequency and the percentage of their pervious performance retained during the probe is reflected in the top panel of Figure 2. These data indicate a positive relation between median training frequencies and the proportion of previous performance retained after a 1-month period without practice. In general, it appears that those participants emitting response frequencies closer to the fluency range on particular skills prior to the retention period (Josie, Larry, & Emily), performed better during the retention probes (over 100% retention) than participants responding at lower frequencies on particular skills. In other words, those participants who rapidly progressed towards fluency on a particular skill during practice, and thus achieved higher overall median training frequencies, appeared to retain more of their performance following the retention period than those who progressed more slowly during training. This relation appears to hold true across participants and, with the exception of Emily, within participants across different basic computation programs.

The relation between each participant's practice opportunities, measured as total responses emitted, and the percentage of their previous performance retained during the probes is reflected in the bottom panel of Figure 2. In contrast to the results obtained for median response frequencies and percentage of retention, these data suggest a negative relation between total responses emitted and percentage of retention across participants. That is, it appears that those participants (Karen and Josie) who engaged in the most cumulative practice (3300 and 4300 responses respectively) on a particular skill retained less of their previous performance. Conversely, those students (Josie and Larry) with the least amount of practice (less than 100 responses) on a particular skill retained as much or more of their performance. This negative relation holds true across participants and, with the exception of Larry and Karen, within participants across the different computation skills. The results for Larry and Karen indicate no relation between amount of practice and skill retention.

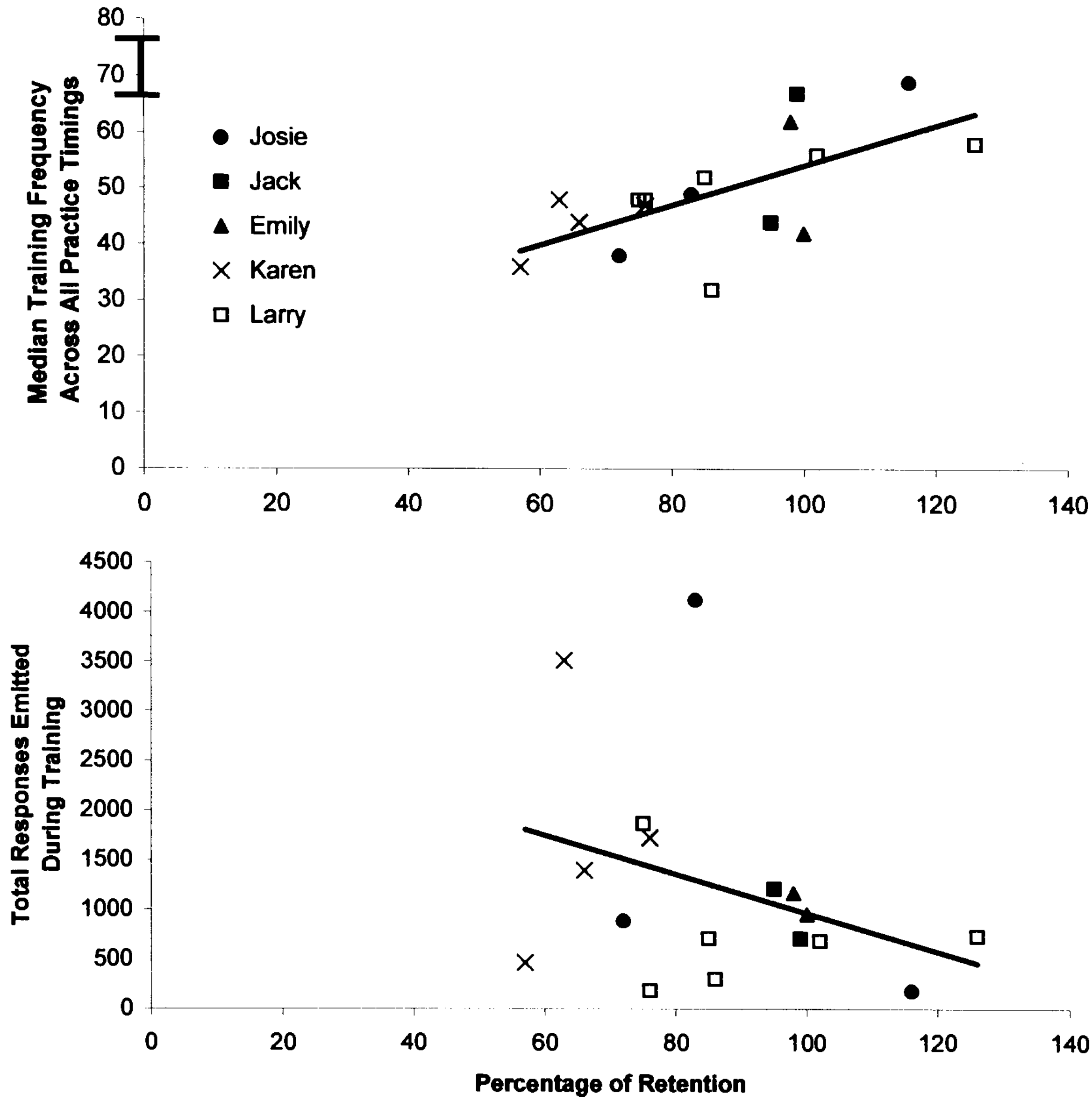
DISCUSSION

The current results indicate a positive relation between response frequencies emitted during practice trials and retention of academic performances by school children. It appears that the higher the frequency of correct responding during practice, the greater the amount of previous performance retained following a 1-month period without instruction or practice. The current findings also indicate a somewhat negative relation between amount of practice (i.e., total responses emitted) and retention of previous performance. It appears that additional practice opportunities will not enhance skill retention unless response frequencies are fairly close to the fluency range. Put differently, engaging in additional practice of a skill at low frequencies did not appear to enhance retention following a 1-month period without practice. These findings suggest that response frequencies emitted during practice, rather than simply the overall amount of practice, is a more critical predictor of skill retention.

There are some limitations to the current study. Firstly, retention probes involved assessing performance on basic computation skills following a 1-month absence from instruction or practice. Future research should focus on an examination of additional skill areas and longer retention periods so that the generality of the current findings can be assessed. Additionally, whether students were actually practicing some skills during the retention period is unknown. Therefore, it is difficult to determine the effect that extraneous practice may have had on the current findings.

The current results contradict previous research that has been conducted investigating relations between practice and retention of academic skills. Overlearning, or exposure to varying numbers of practice trials beyond an initial learning criterion, has been shown to produce retention when measured only as accurate responding (i.e., percent correct). In their metaanalysis, Driskell , Willis, and Copper (1992) indicate that larger de-

Figure 2



Median Training Frequency Across All Practice Timings
80
70
60
50
40
30
20
10
0
Josie
Jack
Emily
Karen
Larry
0
20
40
60
80
100
120
140
Total Responses Emitted During Training
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100
120
140
Percentage of Retention

grees of overlearning yield greater amounts of retention at various follow-up periods. In other words, learners exposed to more additional practice trials during training tend to perform more accurately after a retention period. Thus, a positive correlation appears to exist between amount of practice and skill retention, when retention is measured as accuracy-only. Our results suggest that this relationship does not hold up when frequency is included in the measure of retention.

STUDY 2: ENDURANCE

Method

Participants

According to the criteria described previously, seven students enrolled in CAL were selected for inclusion in the endurance analysis.

Billy (6th grade), Krissi (4th grade), Joey (2nd grade), and Jim (5th grade) were typically developing students receiving instruction in general education classrooms at public or private schools in northern Nevada. Jack (7th grade), Josie (10th grade), and Karen (4th grade), who were included in the retention analysis, were also included in the endurance analysis.

Responses

For all participants, vocally identifying Arabic numerals was included in the endurance analysis. This program requires that students correctly name a series of numbers presented on sheets, with six columns of 16 numbers on each sheet. The numbers presented on each sheet range from the one's to the million's place (depending upon level of difficulty). In other words, the lowest level of the number identification program requires that students name single digit numerals (e.g., 5) and the highest level requires that students name numerals to the millions place (e.g., 1,345,005). Regardless of level, we established the fluency aim for this program according to the recommendation of at least 100 movements (i.e., digits) per minute (Haughton, 1971).

Reinforcers

The reinforcement contingencies were the same as those described for the retention analysis, with one exception. In addition to rewards delivered contingent upon training performance, participants also received highly preferred edible rewards for meeting certain performance requirements during endurance probes. These requirements are described in detail below.

Endurance Probes

Endurance probes were conducted following mastery of a skill level. In other words, when a participant achieved Phase 1 and Phase 2 mastery, his/her endurance across a 5-min timing on that skill was evaluated during the next subsequent session. A 5-min timing length was selected for the endurance analysis so that count per min for each minute of the timing could be plotted on a timings chart and within-timing celeration evaluated. Standard charting convention suggests that at least five data points are required for an accurate measure of celeration (Pennypacker et al., 1972). In addition, relative to 1-min training timings, 5-min timings seemed appropriately lengthy for an evaluation of endurance performance.

Prior to conducting an endurance timing, the tutor informed the participant that he/she would be eligible to choose a prize out of the "endurance box" for engaging in responses per minute that fell at or above the fluency aim (total responses/5). The "endurance box" contained candy bars and other treats much larger in size or value than those items typically earned during sessions. Due to the length of the timing requirement, the aim of these rewards was to motivate participants to try their best during the 5-min timing.

The participant was then presented with five different sheets of stimuli (i.e., Arabic numerals from their current level) placed across the table in a row, with one sheet of stimuli associated with each minute of the timing. Once the timing commenced, the tutor followed along while the participant responded, using identical sheets of stimuli for scoring purposes. At the end of each minute, the tutor prompted the participant to "switch" and respond on the sheet of stimuli associated with the next minute. Throughout the timing, the tutor recorded the correct, incorrect, and skipped response frequencies for each minute. At the end of the timing period, the tutor calculated an overall count per minute for the five minutes (total response frequency/5). If the participant's count per minute for the timing was at or above the fluency aim, then the participant was allowed access to the "endurance box."

Data Analysis

As in the retention analysis, data for the 7 participants were analyzed in terms of median training frequencies and total number of responses emitted during training. The median training frequency and total number of responses emitted were calculated in the same manner as described in the retention analysis. The relation of these measures to frequency per minute across the 5-min endurance timing was explored. The frequency per minute for the 5-min endurance timing was calculated by dividing the total number of responses emitted during the entire timing by five.

Results

The relation between each participant's median training frequency and their frequency per minute during the 5-min endurance probe is reflected in the top panel of Figure 3. These data indicate a positive relation between median training frequencies and frequency of responding during a 5-min endurance probe. In general, it appears that participants emitting response frequencies close to or within the fluency range during training, performed at higher frequencies during the endurance probes. The lowest frequency per minute during the endurance probes was obtained with Karen on a skill where her median training frequency was also below the aim. Josie engaged in the highest median training frequency on a skill and also performed at the highest frequency on that skill during the endurance timing. This relation appears to hold true across participants and, with the exception of Billy and Krissi, within participants across different levels of the numeral identification program.

The relation between each participant's practice opportunities, measured as total responses emitted, and their frequency per minute during a 5-min endurance probe is reflected on the bottom panel of Figure 3. In contrast to the results obtained for median response frequencies and frequency per minute during endurance probes, these data suggest a negative relation between total responses emitted and endurance frequency when results are analyzed across participants. When results are analyzed within participants across the different skill areas there does not appear to be a relation between these two measures. In other words, as with the results for the retention analysis, it does not appear that practice alone is a good predictor of performance across long timing durations. For example, Krissi performed just as well during the endurance probe on a skill with over 3,000 total responses emitted during training as she did on a skill with less than 1,000 total responses emitted during training. Additionally, Karen showed the best endurance performance on a skill with less than 500 total responses emitted during training.

Discussion

The results for the endurance analysis are similar to those obtained for the retention analysis. It appears that participants were better able to maintain stable, high frequencies of correct responses across long timing durations when their median training performance was within the fluency range. It also appears that practice alone was not a good predictor of endurance. In other words, when frequency is not included in the analysis, amount of practice does not predict how well a student will perform over a long timing duration. Engaging in high frequencies of responding during practice appears to enhance a student's endurance, or ability to maintain high correct response frequencies over long timing durations.

Some definitions of endurance include resistance to distractions as well as resistance to fatigue. The current study only evaluated resistance to fatigue through an examination of response stability across a timing duration that was five times as long as the typical practice timing duration. Distractions were not included in the current analysis. However, future research should examine the effects of distracters on performance during endurance timings using frequencies of correct responding as a dependent measure. Future research should also examine performance across increasingly long timing durations as a means of evaluating whether there is a point at which endurance breaks down regardless of median training performance. In this way, educators could better understand how to arrange testing or assessment conditions that encourage a learner's best performance.

STUDY 3: APPLICATION

Method

Participants

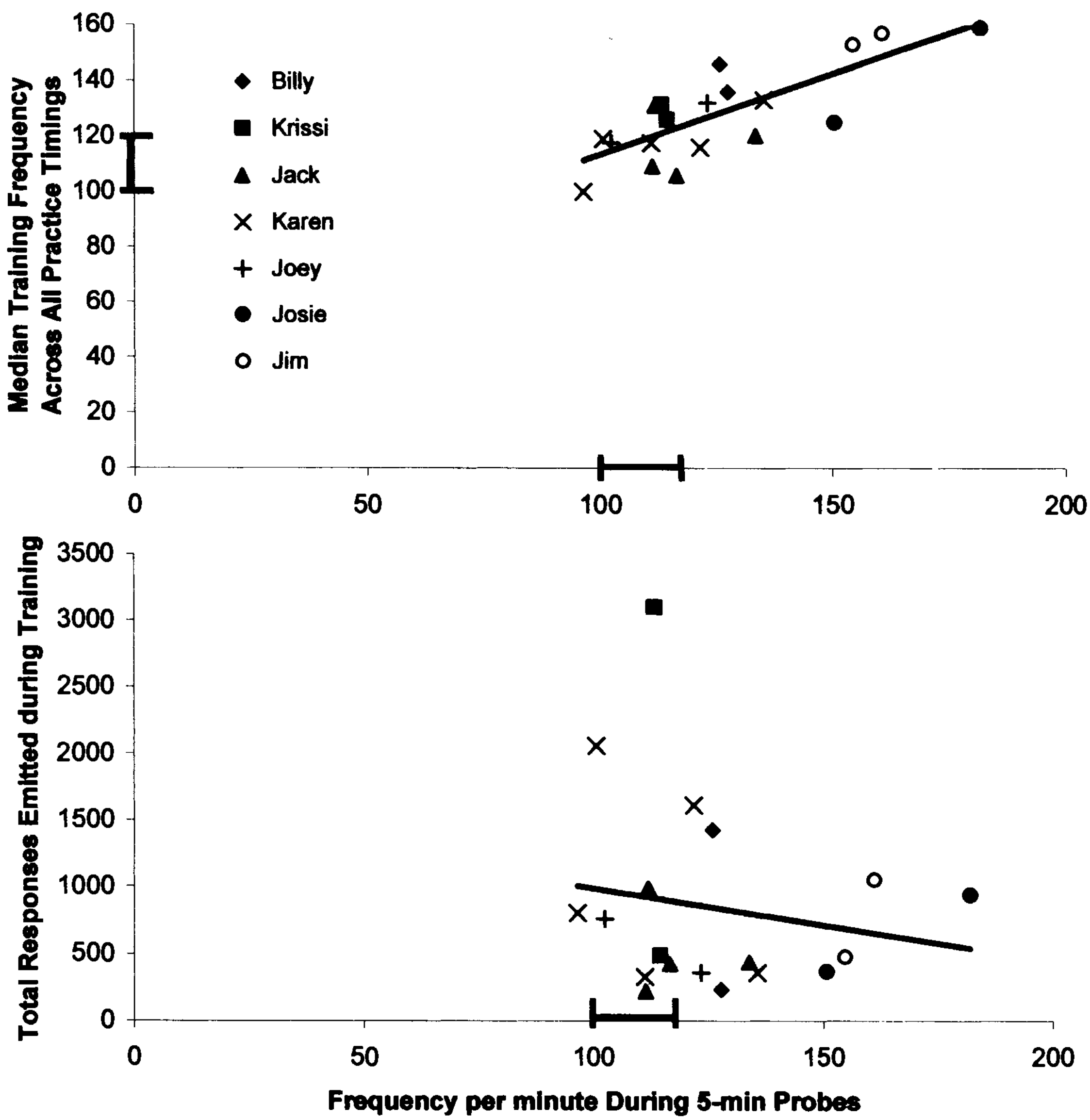
Eight students enrolled in CAL were selected as participants for the application analysis. Participants were selected if they had multiple application probes conducted across multiple skill levels within the composite skill objective defined as Identifying Place Value.

Mike (4th grade), Daryl (4th grade), Emma (4th grade) and Nathan (3rd grade) were typically developing students receiving instruction in general education classrooms at public or private schools in northern Nevada. Billy, Krissi, Joey, and Jim, who were included in the endurance analysis, were also included in the application analysis.

Responses

For all participants, vocally identifying place value was included in the application analysis. This program requires that students correctly

Figure 3



160
140
120
100
80
60
40
20
0
Median Training Frequency
Across All Practice Timings
Billy
Krissi
Jack
Karen
Joey
Josie
Jim
0
50
100
150
200
3500
3000
2500
2000
1500
1000
500
0
Total Responses Emitted during Training
0
50
100
150
200
Frequency per minute During 5-min Probes

identify the place values for a series of numerals presented on sheets, with six columns of 10 numerals on each sheet. On each sheet, one digit in a numeral is printed in a larger font than the others, indicating to the student that they are to name the place value for that digit (e.g., 5, 263 would require a response of "thousands"). The numerals presented on each sheet range from the one's to the million's place (depending upon level of difficulty). In other words, the lowest level requires that students identify numbers in the ones and tens places. The highest level requires that students identify numbers in the ones through one millions places. Regardless of level, we established the fluency aim of at least 90 movements per minute (i.e., correct place values) according to the recommendations of The Haughton Learning Center, who also provided us with the program.

Reinforcers

The reinforcement contingencies were the same as those described in the retention analysis.

Application Probes

Application probes were conducted at the following training milestones: (a) introduction of a skill level and (b) achievement of fluent performance during the various timing lengths (i.e., 15-s, 30-s, and 1-min timings, respectively). During application probes, participants were required to perform the next level of a skill beyond their current training level during a 1-min timing. For example, if a participant's current training level entailed identifying place values through the ten's place, then his/her application probe level entailed identifying place values through the hundred's place. Therefore, when participants reached one of the milestones described above during their current training level, an application probe was conducted at the next subsequent level. Data were collected during application probe timings in the same manner as during training. However, tutors did not provide feedback on or reinforcement for performance during application probes.

Results

Figure 4 depicts response frequencies across application probes for all participants. In general, it appears that response frequencies during application skill levels increased from the initial probe to the final probe. In other words, as participants achieved fluency at the various milestones on skills targeted during training, their performance on higher-level, untargeted skills also increased. There were a few exceptions to this finding. Nathan, Billy, and Jim each showed no change or a decrease in response frequency between two probe timings on a skill. However, for these students, probes were not conducted at every milestone and, therefore, a complete data stream was not available. For those participants where probe data were missing for a particular milestone, fluent performance was obtained at shorter timing lengths when the target skill level was introduced. Thus, frequency building initially commenced during longer timing lengths (i.e., 30-s or 1-min), which prevented application probes from being conducted at earlier milestones. Therefore, some of the data streams depicted on the figure appear incomplete.

Figure 5 depicts the relation between Krissi's training performance on targeted skills and her performance during application probes on untargeted skills. The figure indicates that as Krissi's frequencies of responding on targeted skills increased, her frequencies of responding on untargeted skills during application probe timings increased as well. It is also appears that her time to skill mastery (i.e., fluency) for each subsequent skill level decreased. For example, she required approximately 50 timings to achieve fluency on the first skill level (i.e., 1's-100's), whereas she required only 9 timings to achieve fluency on the fourth level (1's-100,000's). In this way, it appears as though she was able to achieve skill mastery at higher levels more quickly following component skill mastery. This pattern seemed typical when individual performances were analyzed.

Discussion

The current results indicate that increases in response frequencies on targeted skills may lead to increases in frequencies on untargeted skills within the same composite skill objective. In other words, as participants' performances improved on targeted skill levels, corresponding improvements were obtained on higher-level skills prior to the use of specific frequency-building strategies.

These data lend support to the notion of "curriculum leaping," or acquiring upper level skills without direct instruction on those skills (Johnson & Layng, 1992). Given this, requiring more stringent mastery of component skills (i.e., fluency criteria combining accuracy plus speed), might lead to greater overall academic proficiency and critical thinking skills. As many educators have theorized, establishing skills to high frequencies of correct responding might enable those skills to be more readily available for selection by the natural environment. Basic tool skills that occur at high frequencies are more likely to occur on

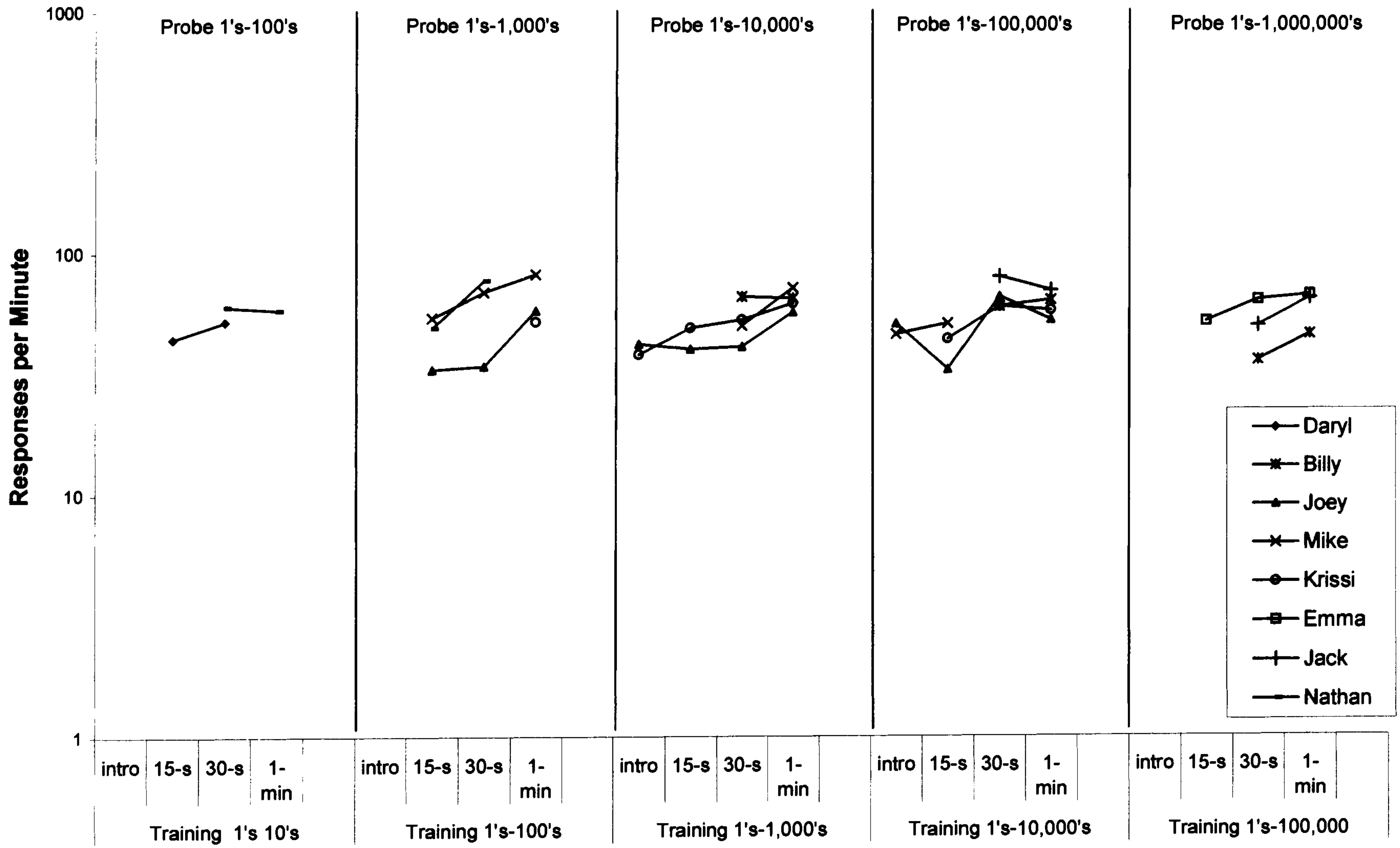

Responses per Minute
1000
100
10
1
Probe 1's-100's
Probe 1's-1,000's
Probe 1's-10,000's
Probe 1's-100,000's
Probe 1's-1,000,000's
Daryl
Billy
Joey
Mike
Krissi
Emma
Jack
Nathan
intro
15-s
30-s
1-min
Training 1's 10's
Training 1's-100's
Training 1's-1,000's
Training 1's-10,000's
Training 1's-100,000

Figure 4

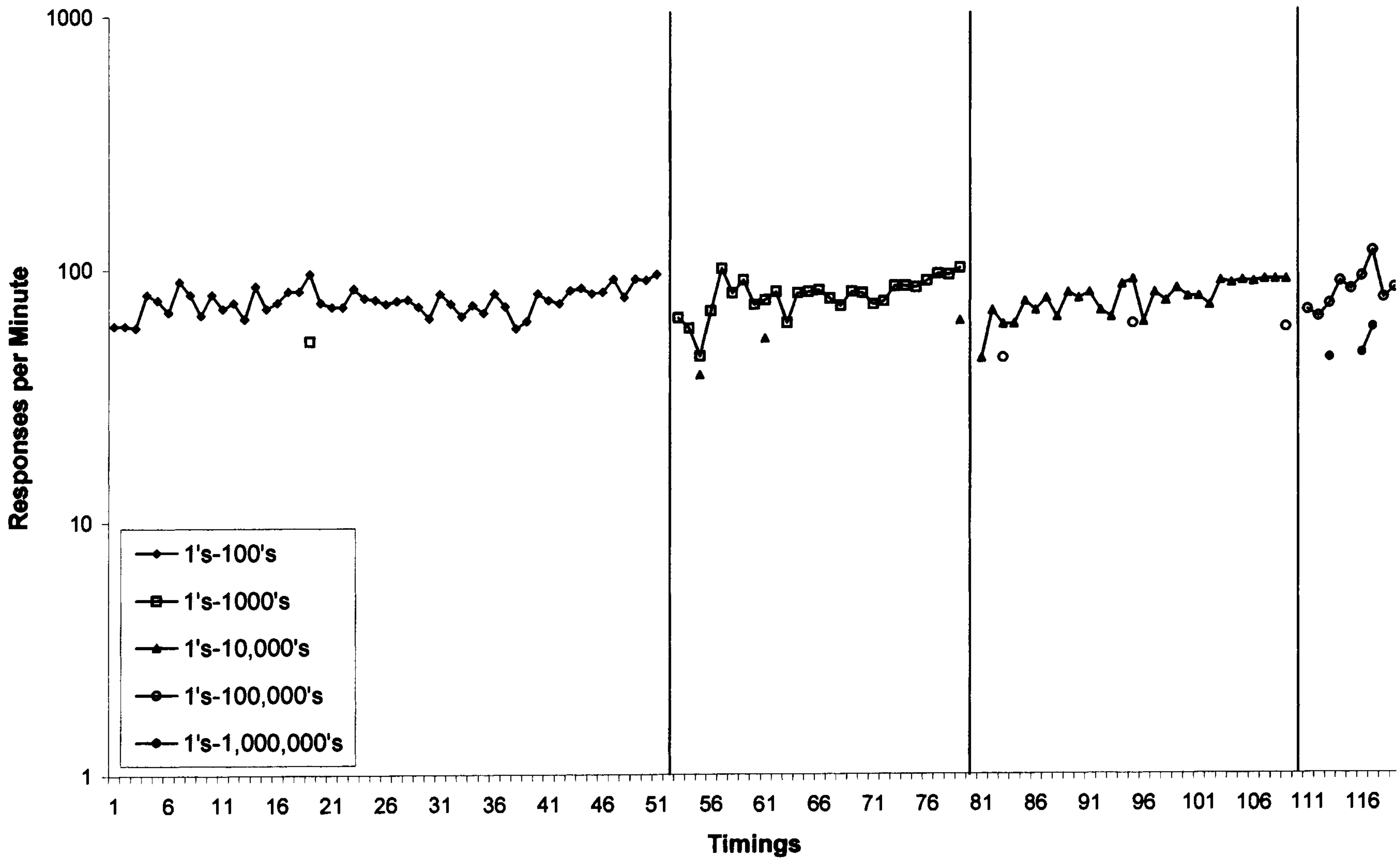

Responses per Minute
1000
100
10
1
Timings
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
106
111
116
1's-100's
1's-1000's
1's-10,000's
1's-100,000's
1's-1,000,000's

Figure 5

novel occasions and recombine with other skills to form novel and more complex academic repertoires (Johnson and Layng, 1992).

The current study serves as a steppingstone for more research in this area. Future research should include analyses of different skill areas across larger numbers of participants as a means of evaluating the generality of the findings. Additionally, unlike the two studies reported previously, practice was not included as a variable in the application analysis. As such, it is difficult to conclude that increases in response frequency on targeted skills were solely responsible for performance improvements on untargeted skills. It may be the case that practice alone improves performance on untargeted skills within the same composite skill objective. Future research examining the role of practice on skill application is needed.

GENERAL DISCUSSION

Overall, the results of the three studies reported have important implications for general education. If frequency of responding is a critical predictor of academic performance outcomes (i.e., REAPS), and speed plus accuracy is a more sensitive measure of academic proficiency than accuracy alone, then fluency-enhancing methods and frequency-based measurement systems within traditional classrooms might lead to greater learner gains. Because accuracy-only measures are most commonly used in education, teachers may be less able to effectively evaluate their students' proficiency in basic skill areas. That is, with an emphasis solely on accuracy, response rate and time required to achieve fluency are overlooked as critical predictors of academic proficiency. As a result, students may be advanced to higher-level skill areas before true mastery is achieved on prerequisite skills. Although some students can perform skills to an accuracy criterion, they may be unable to perform these skills at a rapid pace. This deficit may impede their performance on standardized tests or in more complex skill areas, and lead to academic difficulties at higher-grade levels. The current findings suggest that a more comprehensive picture of academic proficiency requires the inclusion of frequency and accuracy measures.

As mentioned previously, educators adopting CBM methods depend upon the use of frequency measures rather than accuracy measures alone. However, mastery criteria continue to be based upon class norms or averages rather than upon the direct assessment of academic performance outcomes with each learner. In this way, fluency is defined as average performance rather than exemplary performance. Increasing overall academic achievement remains a strong commitment in this country. However, such a goal cannot be achieved when class averages remain the standard in education.

In contrast, PT holds exemplary performance as the standard. However, along with holding students to higher standards of achievement comes the task of devising instructional strategies that will enable all students to reach those standards. As such, educators at the primary grade levels must begin requiring true mastery (i.e., fluency) of basic component skills before allowing advancement to higher-level skills. True mastery must be measured in terms of accuracy plus speed requirements and defined in relation to specific performance standards obtained with individual learners (e.g., REAPS). Additionally, complex skills must be broken down into very basic component units and practiced until true mastery is achieved.

The current results offer a starting point for an effective measurement and analysis tool to use in general education classrooms. Firstly, frequency measures and standard charting practices lead to timely and effective decision-making by educators. Additionally, by collecting frequent (i.e., daily) frequency measures during skill acquisition, educators could compare training performance with other outcome measures so that learning patterns across students and skill areas could be discovered. Thus, more information regarding the instructional conditions that produce the greatest academic proficiency could be identified.

One general limitation of the studies reported above involves each participant's differential progression through the curriculum. For example, some participants had progressed through a significant proportion of the curriculum and thus possessed a large repertoire of basic skills in which fluency had been achieved. As a result, those participants engaged in naturally higher baseline rates of responding during the introduction of novel skill areas than participants who were not as far along in the curriculum. That is, the more experienced students were initially responding at frequencies closer to the fluency range and thus did not have to make as significant an increase to achieve fluency. It may be the case that baseline rates of responding prior to fluency training can predict REAPS. The relationship of base rates to academic performance standards is an area for future research we are currently pursuing.

The current results expand upon clinical and empirical work conducted in the areas of PT and CBM that emphasize direct measurement of behavior (i.e., frequency) as a basic assessment tool. In order for educators to better understand the conditions necessary for establishing and maintaining overall academic proficiency, more research needs to be conducted in the areas of PT and flu-ency-based instruction. With more frequent and rigorous investigation, it may be possible to identify educational technologies where academic excellence is the norm rather than the exception. We hope that these studies will set the occasion for more research in this area.

REFERENCES

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19, 163-197.

Bloom, B.S. (1986). The hands and feet of genius: automaticity. *Educational Leadership*, Feburary, 70-77.

Deno, S. (1985). Curriculum-Based Measurement: The emerging alternative. *Exceptional Children*, 52, 219-231.

Driskell, J.E., Copper, C., Willis, R.P. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, 77, 615-622.

Elliot, S.N. & Fuchs, L.S. (1997). The utility of Curriculum-Based Measurement and Performance Assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review*, 26, 224-233.

Fuchs, L.S. & Deno, S. (1992). Effects of curriculum within Curriculum-Based Measurement. *Exceptional Children*, 58, 232-243.

Fuchs, L.S., Deno, S., & Mirkin, P.K. (1984). The effects of frequent Curriculum-Based Measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, 21, 449-460.

Haughton, E. (1971). Aims: Growing and Sharing. In J.B. Jordan & L.S. Robbins (Eds.), *Let's try doing something else kind of thing* (pp. 20-40). Arlington, VA: Council on Exceptional Children.

Hintze, J.M., Conte, K.L., Shapiro, E.S., & Basile, I.M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review*, 26, 535-553.

Johnson, K.R., & Layng, T.V.J. (1992). Breaking the structuralist barrier. Literacy and numeracy with fluency. *American Psychologist*, 47, 1475-1490.

Lindsley, O.R. (1971). From Skinner to precision teaching: The child knows best. In J.B. Jordan & L.S. Robbins (Eds.), Let's try doing something else kind of thing (pp. 1-11). Arlington, VA: Council on Exceptional Children.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-Based Measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education*, 18, 109-117.

Pennypacker, H.S., Koenig, C.H., Lindsley, O.R. (1972). *Handbook of the Standard Behavior Chart*, Kansas City, KS: Precision Media.

Wesson, C., Skiba, R., Sevcik, B., King, R.P., & Deno, S. (1984). The effects of technically adequate instructional data on achievement. *Remedial and Special Education*, 5, 17-22.

Wilson, Schendal, & Ulman (1992). Curriculum-based measures, teachers' ratings, and group achievement scores: Alternative screening measures. *Journal of School Psychology*, 30, 59-76.