EFFECTS OF CELERATION RATE ON BEHAVIORAL FLUENCY

By

Kevin J. Munson, M.A.

A DISSERTATION

Submitted to
The Eberly College of Arts and Sciences
at
West Virginia University

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Behavior Analysis

Department of Psychology

Morgantown, West Virginia
1998
ACKNOWLEDGEMENTS

I would like to acknowledge John Crosbie for dedicating so much of his time and effort to my training as a behavior analyst, and for always displaying the highest standards of professionalism. I owe a great debt of gratitude for everything John has done for me throughout my time at West Virginia University (WVU). I also would like to recognize the contributions of my fellow graduate students, including Leo Carlin, Dave Ellenwood, Janet Emmendorfer, Kristin Hazlett, Mike Layng, Greg Madden, Finnur Oddsson, and Oliver Wirth, who provided many ideas related to how fluency should be studied in the laboratory. I would specifically like to thank Leo Carlin for helping me design the celeration contingencies and programming algorithms while we watched Monday-night football games. I also would like to recognize the helpful suggestions provided by the members of my dissertation committee, including Phil Chase, John Crosbie, Don Kincaid, Mike Perone, and Anne Watson. Finally, I would like to thank Darnell Lattal for encouraging my professional development in so many ways.

The research reported in this dissertation was funded by grants from (a) the WVU Department of Psychology Alumni Fund, (b) the WVU Eberly College of Arts and Sciences, (c) the Office of Academic Affairs at WVU, and (d) the Sigma Xi Scientific Research Society.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS.......................................................................................................... ii

TABLE OF CONTENTS............................................................................................................. iii

LIST OF TABLES.................................................................................................................... v

LIST OF FIGURES ................................................................................................................ vi

INTRODUCTION .................................................................................................................. 1

   Behavioral Education........................................................................................................... 2

   Fluency ................................................................................................................................. 6

   Contingency Adduction ....................................................................................................... 11

   Behavior Streamlining ....................................................................................................... 13

   Automaticity ....................................................................................................................... 14

   Experimental Question ..................................................................................................... 15

GENERAL METHOD ............................................................................................................... 17

   Subjects and Setting.......................................................................................................... 17

   Apparatus ......................................................................................................................... 18

   General Procedure ............................................................................................................ 20

   Reinforcement Contingencies ......................................................................................... 24

   Tests for REAPS Outcomes .......................................................................................... 28

EXPERIMENT 1: Fluency Aims ............................................................................................ 33

   Method ............................................................................................................................... 33

   Procedure ......................................................................................................................... 33

   Results and Discussion ................................................................................................. 34

   Effectiveness .................................................................................................................. 34
Efficiency ..................................................................................................................... 37

EXPERIMENT 2: Celeration ..................................................................................................... 51
Method .............................................................................................................................. 51
Procedure .......................................................................................................................... 51
Results and Discussion ........................................................................................................ 51

GENERAL DISCUSSION ......................................................................................................... 72
Interpretation of Results ........................................................................................................ 72
Topography ........................................................................................................................ 78
Experimental Procedures .................................................................................................. 80
Voice Recognition Apparatus .............................................................................................. 81
Future Studies ..................................................................................................................... 82
Applied and Social Validity ................................................................................................. 84
Conclusions .......................................................................................................................... 85

REFERENCES ..................................................................................................................... 87

APPENDIX A: Subject Recruitment Poster ........................................................................... 94
APPENDIX B: Programmed Instruction Sequence ................................................................. 95
APPENDIX C: Utterance Duration and Amplitude ................................................................. 99

ABSTRACT .......................................................................................................................... 107

APPROVAL OF THE EXAMINING COMMITTEE ................................................................. 108
LIST OF TABLES

Table 1: Example Target IRTs with Low and High IRT Range Limits........................................ 26

Table 2: Summary of Tonal Frequency and Duration, Textual Feedback Messages, and Reinforcement Conditions ........................................................................................................ 27

Table 3: Experiment 1: Summary of REAPS Posttest Data ....................................................... 35

Table 4: Experiment 2: Scores on REAPS Posttests ............................................................... 62

Table 5: Experiment 2: Practice Trials, Reinforcers per 1-hr session (Rein), and Percent Correct During Fluency Building................................................................. 64
LIST OF FIGURES

Figure 1. Sample feedback chart shown to subjects between fluency-building timings .......... 24

Figure 2. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 25 responses per minute rate criterion group ..................................... 38

Figure 3. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 30 responses per minute rate criterion group ..................................... 39

Figure 4. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 35 responses per minute rate criterion group ..................................... 40

Figure 5. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 40 responses per minute rate criterion group ..................................... 41

Figure 6. Experiment 1: Percent of responses during 1-min timings with IRTs within the range of IRTs eligible for reinforcement for the 30 responses per minute group .............. 44

Figure 7. Experiment 1: Percent of responses during 1-min timings with IRTs within the range of IRTs eligible for reinforcement for the 35 responses per minute group .......... 45

Figure 8. Experiment 1: Percentage of reinforcers earned per available opportunity for subjects in the 30 responses per minute group ................................................................. 46

Figure 9. Experiment 1: Percentage of reinforcers earned per available opportunity for subjects in the 35 responses per minute group ................................................................. 47

Figure 10. Experiment 1: Performance on probe tests for the 30 responses per minute group ... 50

Figure 11. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x2 celeration group ................................................................. 53

Figure 12. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x2 celeration group ................................................................. 54
Figure 13. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x2 celeration group.......................... 55

Figure 14. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x2 celeration group.......................... 56

Figure 15. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x1.25 celeration group........................................................................... 58

Figure 16. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x1.25 celeration group........................................................................... 59

Figure 17. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x1.25 celeration group.......................... 60

Figure 18. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x1.25 celeration group.......................... 61

Figure 19. Experiment 2: Median latency and inter-quartile ranges across 1-min timings for subjects in the x2 celeration group........................................................................... 66

Figure 20. Experiment 2: Median latency and inter-quartile ranges across 1-min timings for subjects in the x2 celeration group........................................................................... 67

Figure 21. Experiment 2: Median latency and inter-quartile ranges across 1-min timings for subjects in the x1.25 celeration group........................................................................... 68

Figure 22. Experiment 2: Median latency and inter-quartile ranges across 1-min timings for subjects in the x1.25 celeration group........................................................................... 69

Figure 23. Medians and inter-quartile ranges for utterance duration and utterance amplitude for two representative subjects from the x1.25 and x2 celeration groups................. 71
INTRODUCTION

Many instructional techniques used in behavioral education are based on discoveries from the experimental behavior laboratory. For example, programmed instruction (Skinner, 1954) uses reinforcement, shaping, prompting, and fading. Occasionally, behavioral education contributes new principles to be studied by experimental behavior analysts. One such principle is behavioral fluency.

Behavioral fluency is functionally defined as the rate of responding that predicts that (a) the response can be performed accurately and rapidly even after a delay since its acquisition, (b) performance of the response can be maintained for long intervals, (c) the response will facilitate learning more complex responses, and (d) rates of responding can be maintained during distracting conditions (Johnson & Layng, 1992, 1994, 1996). This definition provides objective outcome-based criteria that can be used to assess whether a particular response is fluent.

Unfortunately, it is difficult to determine rates necessary to obtain fluency on any particular skill. Furthermore, there have been only a few scientific investigations of the variables that produce fluency. Nonetheless, behavioral-education practitioners continue to employ fluency procedures in the classroom without an experimental analysis of critical variables. For example, it is assumed that the optimal method for achieving fluency is to increase the rate of a response rapidly until it reaches a level (i.e., fluency aim) associated with the learning outcomes mentioned previously. This assumption, however, has not been tested rigorously (Binder, 1996).

For many years, the rate at which fluency aims are attained has been proposed as a critical variable in the development of fluent performance (Johnson & Layng, 1992). Recently, however, some people have questioned whether it is necessary for responding to reach high rates rapidly to produce fluent performance (Binder, Lindsley, Johnson, Parsons, Eshleman, &
Pennypacker, 1995; Munson & Crosbie, 1998). Behavioral educators have asked this question, but are looking to laboratory researchers for the answer (Binder, 1996; Carlin, Wirth, Munson, & Chase, 1996; Lindsley, 1996a). The present studies addressed this question, and were conducted in a laboratory setting to maximize experimental control.

In this introduction, there are three main sections: First, the historical development of fluency is discussed in the context of approaches to behavioral education. Second, empirical criteria for evaluating fluent performance, relevant research on fluency, and key elements of fluency procedures are summarized. Third, psychological and learning concepts related to fluency are explored including (a) contingency adduction, (b) behavior streamlining, and (c) automaticity.

**Behavioral Education**

From a behavioral perspective, education is defined as the process by which the behavior of learners (students) is modified as a result of being exposed to a structured sequence of instructional materials (Johnson & Ruskin, 1977; Pennypacker & Binder, 1992). Inherent to most behavioral education techniques is a focus on (a) objectively defining the behavior to be learned, (b) designing stimulus materials to teach that behavior based on empirically-validated hierarchies or sequences of instructional objectives (Robin, 1976), (c) delivering the materials to learners, (d) modifying techniques based on individual responses, and (e) individualization of learning experiences to suit each student and maximize learning outcomes (Anderson & Faust, 1973; Barrett, et al., 1991).

The predominant behavioral education techniques are programmed instruction (Skinner, 1954), personalized system of instruction (PSI; Keller, 1969), direct instruction (Becker & Carnine, 1980; Carnine & Silbert, 1979; Engelmann & Bruner, 1988; Kinder & Carnine, 1991;
Silbert, Carnine, & Stein, 1981), precision teaching (Lindsley, 1971a), and generative instruction (Johnson & Layng, 1992). The construct of behavioral fluency comes from precision teaching and generative instruction (Binder, 1996; Binder & Bloom, 1989). Consequently, subsequent sections focus on relevant features of these two approaches.

**Precision teaching.** Lindsley developed the techniques of precision teaching when he was studying behavior of people with mental illnesses in the laboratory under the supervision of Skinner in the 1950’s (Lindsley, 1990a). Lindsley expressed dissatisfaction with educational approaches of that time because they assessed learning with percent correct, which, he argued, imposes an artificial ceiling on measurement of learning. That is, when response accuracy is 100%, no additional learning can be measured. Instead, Lindsley argued that learning should be measured with response rate (Lindsley, 1996a).

Lindsley claimed, for example, that a student who could complete 10 mathematics problems per minute, even at 100% accuracy, was less capable at math than a student who could complete 50 per minute at 100% accuracy (1971b). Hence, the fundamental assumption of precision teaching is that ongoing rate of academic behavior is a better indicator of learning than is percent correct. Frequency of responding is of primary interest to precision teachers because of its sensitivity to educational interventions, and because it focuses on response probability rather than accuracy alone.

Precision teaching is not a teaching method. Rather, it is a set of measurement and assessment procedures to be used in combination with any other teaching method (Koenig, 1972; Neal, 1981; West, Young, & Spooner, 1990). Critical components of the precision-teaching model are (a) use of functional definitions of behavior (Downs & Morin, 1990), (b) use of response rate as a measure of academic behavior, (c) measurement of student learning by
frequent performance assessment (Beck & Clement, 1991), (d) frequent graphing of student behavior on a standardized chart (Lindsley, 1971a; Pennypacker, Koenig, & Lindsley, 1972), (e) frequent data-based analysis of the impact of teaching interventions (Potts, Eshleman, & Cooper, 1993), and (f) emphasis on building socially important behavior (Lovitt et al., 1990; Neal, 1981; West et al., 1990; White, 1986). Precision teaching allows any teacher to describe academic behavior precisely, deliver instruction designed to teach that behavior to students, record the results of teaching, and use specific decision rules to intervene (Beck & Clement, 1991).

Precision teachers use celeration as the primary indicator of learning progress (Brandstetter & Merz, 1978). For the purposes of the present experiments, celeration refers to the “trend in a time series of [response] frequencies expressed as a multiplication… in frequency per week of calendar time (Binder 1996, p. 167)”. A celeration target of x2 (read as “times two”) per week means that response rate doubles each week.

The current practice in precision teaching is to provide an average of twenty 1-min timings per week (J. W. Eshleman, personal communication, April 6, 1997). To maintain a celeration target of x2, therefore, a student who responds at 10 responses per minute must respond at 20 responses per minute one week (and 20 timings) later. Similarly, if response rate increases by 25% each week, then celeration is x1.25. Precision teachers use x2 as an optimal celeration rate for learning, and x1.25 as a minimum acceptable celeration rate (Binder, 1996). In the present experiments, maximum and minimum celeration rates were x2 and x1.25, respectively.

**Generative instruction.** The behavioral-education technique known as generative instruction was developed by Johnson at the Morningside Academy in Seattle and Layng at
Malcolm X College in Chicago (Binder & Johnson, 1991; Johnson & Layng, 1992, 1994). It is called generative because the instructional procedures are designed to generate complex academic behavior with little or no training (Alessi, 1987). It is not a new educational technique, but rather a novel combination of features from other behavioral-education approaches.

Generative instruction borrows response rate as a dependent measure of learning from precision teaching, repertoire-establishing techniques and instructional content analysis (Markle, 1967) from programmed instruction, structured curriculum and instructional analysis (Engelmann & Carnine, 1982) from direct instruction, self-pacing from personalized system of instruction (Keller, 1969), and simulation exercises from computer-assisted instruction (Johnson & Layng, 1992).

Progression through a generative-instruction curriculum is based on (a) individualized placement of each student that relies upon extensive normative testing and interviews; (b) clear statements of learning objectives; (c) carefully sequenced instructional materials, based on learning objectives that are broken into a series of small steps; (d) student self-pacing; (e) immediate feedback on performance; (f) individual, peer, and small-group learning; (g) practice techniques based on precision teaching, (h) monitoring of learning until mastery of each behavioral objective is obtained; and (i) reinforcement of attendance and rapid progress (Binder & Johnson, 1991; Johnson & Layng, 1992, 1994, 1996).

Although there have been almost no controlled studies of generative instruction (Robbins, 1994), anecdotal reports suggest that it is effective. At Morningside Academy, for example, students typically gain between two and three grade levels on standardized tests in each academic skill each year (Johnson & Layng, 1992, 1994).
Generative instruction has many features in common with precision teaching, such as a focus on response rate as the primary indicator of learning, timed practice sessions, and graphing of student performance. In fact, many precision teachers are trained in the methods of generative instruction (Binder et al., 1995). It is from these behavioral education domains that behavioral fluency emerged.

**Fluency**

Precision teachers capitalized on the finding that students’ performance on component skills predicts learning on other higher-level skills (Binder, 1993). The term fluency seemed to tact the general response topography that allowed learners to achieve performance gains at extremely high rates of responding. Qualitative differences between experts and amateurs are directly related to frequency of responding (McDade & Goggans, 1993). For example, it is estimated that few individuals have ever attained the status of international chess grandmaster without at least 10 years of practice playing chess, and that most chess moves are made by such experts in approximately one-third of the time taken by amateurs (Ericsson, Krampe, & Tesch-Romer, 1993).

Similarly, precision teachers view response frequency (i.e., rate) as a parameter of academic behavior that directly predicts accelerated student learning. In fact, it has been suggested that the more rapidly that learners’ responding reaches a high rate, the easier future learning will become (Binder et al., 1995). That is, not only is final response speed important, but so is the speed with which learners reach that high speed. This proposition was tested in the present research.

**REAPS criteria.** Precision teaching and generative instruction sought to define fluency according to functional outcome criteria. According to some definitions, fluency is merely
accuracy plus speed of performance (e.g., Binder, 1988, 1990, 1991, 1993). This definition lacks descriptive power, because it says nothing about the behavioral outcomes associated with performance that is both fast and accurate. Furthermore, it borders on circularity.

The original acronym used to describe fluency outcomes, RAPS, came from Haughton, a graduate student working with Lindsley in precision teaching (Lindsley, 1996a). Haughton’s RAPS stood for retention, application, and performance standards: three putative outcomes related to fluent performance. Lindsley later added E for endurance, and converted the S to stability (Johnson & Layng, 1996), which refers to a behaviors’ minimal change in rate during distraction conditions (Lindsley, 1996b). Hence, the acronym REAPS (Retention, Endurance, Application, Performance standards, and Stability) is used as a set of functional criteria to assess whether a particular behavior is fluent. Unfortunately, the value of using the REAPS criteria to describe fluency has not been tested experimentally. The present experiments provided such a test.

Functional definitions of REAPS criteria have been refined recently. Retention refers to maintenance of the rate of a particular skill between two measurement periods, between which the learner has not been required to perform the skill (Binder, 1996). Measurement intervals greater than two weeks are usually used to assess retention (Johnson & Layng, 1994). Endurance is demonstrated when the learner can perform a skill while maintaining a high rate for a longer measurement interval than was used during training. The typical measurement duration used to assess endurance is three or more times the interval used during training (Haughton, 1977). Application means that the skill can be used as a component of higher-order skills. For example, a student who is fluent at simple math calculations such as addition and subtraction can apply that skill when solving algebraic equations. Performance standards refer to the ability of a
learner to perform at a rate that predicts the occurrence of other learning outcomes associated with fluency (i.e., retention, endurance, application, and stability). Stability means that a particular rate of performance on a task will be maintained when that task is subjected to distracting conditions. For example, subjects who were taught to read Hebrew characters at high rates maintained those rates when random numbers were read to them through headphones, but subjects who did not read Hebrew at high rates were easily distracted by hearing random numbers (Binder 1988, 1993).

Performance standards used in the classroom (also known as fluency or frequency “aims”) are based on predetermined rates of performance that predict REAPS outcomes will occur. One practical consideration for Precision Teachers developing REAPS standards is that performance frequency aims need to be developed for diverse individual academic skills. This process is reputedly an empirical one that is based on years of observation of students in precision teaching classrooms (Binder, 1993).

A variety of techniques are used to set frequency aims, most of which are based on normative criteria. For example, Wood, Burke, Kunzelman, and Koenig (1989) developed aims for various math skills by identifying adults who use math in their occupation, sampling the frequency of their correct answers per minute on relevant math skills, then used data from these adults to set aims for students. Other techniques include (a) peer comparison (Evans & Evans, 1985); (b) establishing aims based on previous performance on a related skill (Neal, 1981); (c) use of an adult-to-child proportional formula (Binder, 1996); (d) functional rates (i.e., rates that predict successful progression through curricula; Brent, 1977; Johnson & Layng, 1996); and (e) direct measurement of the performance of target groups of individuals who are recognized as experts in the skill domain being measured (Howell & Lorson-Howell, 1990; Pennypacker &
Binder, 1992). These techniques sometimes are used to circumvent the requirement that rate standards should be empirically developed for each REAPS criterion for each class of academic behavior (Binder, 1996), which may not be feasible.

For the purposes of experimental analysis, an uncomplicated strategy for determining rate criteria (i.e., fluency aims) for a particular skill is to find the minimum rate that predicts that subjects can retain, endure, and apply the skill, and perform it without being distracted. “Simply demonstrating in a systematic fashion that higher performance frequencies improve outcomes in one or more of the REAPS categories for one behavior class is itself a notable accomplishment, one that can surely inspire many theses and dissertations in the future (Binder, 1996, p. 174)”.

In accordance with Binder’s suggestion, rate criteria were determined experimentally in the present experiments.

**Studies of fluency.** The standard form of professional communication in precision teaching and generative instruction has been informal sharing of behavior charts at conferences and educational seminars, not publication in peer-reviewed journals (Binder, 1996). A large amount of data gathered in precision-teaching classrooms is stored in a computer database maintained by Lindsley called the “Behavior Bank” (Koenig, 1971). The Behavior Bank was originally designed to be a source of data on response frequency that any precision-teaching researcher or practitioner could access. Unfortunately, researchers have not used it.

Only a few experimental studies of fluency have been reported in peer-reviewed journals (Robbins, 1994). For example, Evans and Evans (1985) studied the effects of training math skills and consonant-vowel-consonant sounds to various rates on the subsequent acquisition of more complex algebra and reading skills. Subjects were trained to low, medium, and high rates on component skills, then tested on subsequent tasks in a curriculum. It was found that higher
rates generally facilitated acquisition of more complex skills, although the performance of subjects who performed at a moderate rate was not significantly different from that of subjects trained to high rates.

Ivarie (1986) attempted to determine the effects of two frequency aims on retention. Roman numerals were selected for the experimental task because subjects had little exposure to them. Subjects in average and below average academic categories performed significantly higher three months after training when a high response rate was required. In addition to these studies, applied research conducted by precision teachers generally supports the usefulness of building skills to high rates (Beck & Clement, 1991; Binder & Watkins, 1990; Reynolds & Glaser, 1964; Spiegel & Bryant, 1978; Van Houten & Rolider, 1989; Weinstein & Cooke, 1992; Yaber & Malott, 1993).

**Fluency building.** Johnson and Layng (1996) recommend that experimental assessments of fluency should adopt current practices in precision teaching and generative instruction to ensure that subsequent findings are relevant to current practice. Incorporating every aspect of these teaching models, however, is neither feasible nor desirable in an experimental setting, because that would introduce too many uncontrolled factors. Hence, the present research focused mainly on a particular stage of the precision teaching process - fluency building.

Fluency building typically occurs following direct-instruction or programmed-instruction training that increases skill accuracy to 85% correct or greater (Howell & Lorson-Howell, 1990). Fluency building involves a series of 1-min timed practice sessions in which students respond to stimulus materials on a printed page or computer screen. Practice stimuli usually are arranged such that there are more trials than students can possibly complete in one practice
timing (Binder, 1996; White & Haring, 1976). This is done so that rate is not constrained by the number of practice items.

Fluency coaching – a technique in which reinforcing consequences from teachers or other students are presented rapidly during practice sessions – is used to increase response rate (Binder, 1996). Correct and incorrect responses are counted following each practice session, and students display those data on charts. A minimum of 30 s usually elapses between successive timings (Binder, Haughton, & Van Eyk, 1990). Students continue practicing until their rate of correct responses meets or exceeds a fluency criterion (i.e., a predetermined rate that is assumed to predict good performance on REAPS outcomes). In general, most basic skills (e.g., math facts such as addition, subtraction, multiplication, and division) become fluent at approximately 40 to 60 responses per minute, although the most elementary skills (e.g., writing alphanumeric characters) don’t become fluent until approximately 200 responses per minute (Lindsley, 1996b). In this regard, response rates necessary for fluency vary greatly depending on the difficulty of educational tasks.

Students are not allowed to progress beyond the fluency-building stage until they can perform at or above frequency aims while maintaining 90-95% accuracy levels (Howell & Lorson-Howell, 1990). Students reach frequency criteria more rapidly if they practice tasks with fewer skills in the set (Evans & Evans, 1985). In the present experiments, the following typical fluency-building procedures were used: (a) 1-min practice timings, (b) 30-s breaks between timings, (c) fluency and accuracy criteria, and (d) a limited set of instructional stimuli.

Contingency Adduction

It has been argued (Binder, 1996) that fluency building allows basic skills (components) to be adduced by instructional contingencies present in higher-order skills (composites). That is,
two or more component skills that initially are learned to fluent levels under separate stimulus conditions can be combined easily into a new composite skill if the stimulus conditions are arranged to evoke such a response. Johnson and Layng (1994) call this phenomenon contingency adduction because it involves the recombination of existing fluent behaviors into novel repertoires (Andronis, 1983).

Behavioral educators have shown that mastery of lower objectives in a hierarchy of skills facilitates learning of higher objectives (Birch, 1945; Haughton, 1972; Liberty, Haring, White, & Billingsley, 1988; Resnick, Wang, & Kaplan, 1973). Fluency may expedite this process by strengthening the probability of component responses. Fluent component repertoires may combine to produce new sequences or chains of behavior, behaviors with different functions, or behaviors with different topographies.

When individual topographies are compatible, it may be more likely that contingencies will adduce multiple responses (Epstein, 1985; Epstein & Medalie, 1983), thereby producing novel behavior. For example, an artist who is fluent at two different brush strokes may combine the strokes to produce a desired visual effect that could not have been produced with either brush stroke alone. This may occur although the specific combination of brush strokes was never explicitly trained.

Recently, Johnson and Layng (1996) suggested appending contingency adduction to the existing REAPS fluency criteria and changing the acronym to RESAA (retention, endurance, stability, application, and adduction). Although Johnson and Layng gave no specific functional guidelines for how to test for adduction, in the present experiments this was done by testing untrained higher-order composite skills after training component skills to fluent levels.
Adduction was not used as a fluency criterion in the present experiments, however, because its conceptual status is currently unknown.

**Behavior Streamlining**

Behavior streamlining is a theoretical process related to fluency building. It has been proposed that the high rates necessary to produce fluent performance “streamline” behavior by selecting the most efficient response topographies (K.R. Johnson, personal communication, August 10, 1996). For example, the topography of dysfluent counting behavior may include using one’s fingers to count, whereas the topography of fluent counting does not. Using fingers to aid in counting does not allow learners to achieve high rates; therefore, this topography is selected out by instructional contingencies (cf. Johnson & Layng, 1992).

Dramatic changes in response topography during fluency building may be correlated with changes in learning that accompany fluency (i.e., REAPS outcomes). When streamlined topographies of different skill repertoires are similar in form, instructional contingencies may be more likely to select a novel combination of these repertoires, which could then be called adduction.

Although behavior streamlining is a new proposition that has no empirical support, the present experiments provided data that may shed light on the phenomenon. Specifically, response latency, duration, and amplitude of vocal responses were measured to assess whether changes in response topography occur during fluency building. Reading binary numbers may involve covert cognitive behavior topographies, such as counting place values and converting digits. Although cognitive behavior cannot be measured by available techniques, it is possible that topographical measures of vocal responses are indicators of covert behavior, so they were recorded.
Automaticity

Findings from related areas of research may shed light on some of the phenomena associated with fluency. Fluency is closely related to what cognitive psychologists call “automaticity” (Binder, 1991; Samuels, 1987). Skills are automatic when subjects can engage in them at high rates without conscious attention (Bloom, 1986), and can do many things concurrently.

Automaticity is studied with discrete-trial procedures, where subjects repeat a behavior for varying amounts of practice until response latencies (i.e., reaction times) decrease to below a threshold level. Subjects are then given tasks that require use of the practiced skill under distraction conditions, and the skill is said to be automatic when distraction conditions are not correlated with increases in latency from automatic levels. Automaticity also is associated with increases in accuracy (Bloom, 1986), retention, and application (Binder, 1996).

The conceptual rationale for why automaticity produces these effects is quite different from anything discussed in the fluency literature (Dougherty & Johnston, 1996). Cognitive psychologists maintain that mastery of any skill depends on the ability to perform it with both speed and accuracy (i.e., with fluency), while continuing other brain functions (Bloom, 1986). To develop automaticity in any skill requires identification and consistent practice of both the skill and sub-skills that may be necessary to perform it. As responding becomes automatic, it uses fewer cognitive resources, thereby allowing skills to be performed unconsciously.

Behavior analysts have noted that there are similarities between fluency and automaticity, but conceptual and methodological differences (e.g., free-operant versus discrete-trial procedures) make it difficult to synthesize the two concepts (Robbins, 1994). Automatic behavior appears to be stereotypical and rapid, as if subjects could not withhold the response
given the appropriate stimulus conditions. Perhaps changes in topography of skills when they become automatic processes are tantamount to topographical changes in fluent responding (i.e., behavior streamlining). Although the conceptual relation is unclear, it seems prudent to measure changes in response topography of fluent behavior to see if stereotyped response topographies are correlated with increases in accuracy or the ability to apply the skill. For these reasons, response topography (i.e., latency, duration, and amplitude of vocal responses) was recorded in the present experiments. If these topographical measures change dramatically as responding becomes fluent, it might be an indicator that fluency selects the most efficient response topography.

**Experimental Question**

Recently, Johnson and Layng (1996) noted that “it is time to develop a coherent investigation of fluency (p. 281)” In my opinion, a proper investigation of fluency should begin by examining the various components of fluency that may be critical to its development. Precision teaching and generative instruction both assume that rate of response is the critical factor in building fluent behavior (Binder, 1996; Johnson & Layng, 1996; Lindsley, 1996a), yet it is unknown whether celeration also may affect fluency outcomes. That question was addressed in the present study.

The present experiments were designed to answer one primary question: Is behavior that reaches a particular rate rapidly functionally different from behavior that takes longer to reach the same rate? The principal dependent variables were performance on REAPS tests and response topography; independent variables were terminal rate and celeration rate.

To answer the experimental question, it was necessary to establish an appropriate terminal rate that predicts fluent performance on the experimental task. The goal of
Experiment 1 was to find the appropriate fluency aim for the experimental task. A frequency criterion was selected for use in Experiment 2 by assessing the effects of different terminal response rates on performance on REAPS outcomes. In Experiment 2 that frequency criterion was used to test the effects of two celeration rates \((x2 \text{ and } x1.25)\) on fluency outcomes.

A task that was unfamiliar to subjects was used, plus advanced voice-recognition technology to measure various topographical aspects of responding. The experimental task was reading numbers from the binary number system (i.e., Base 2) in the decimal range of 0 to 31.

A computer was used to present instructional materials and measure responses precisely. Binder (1988) maintains that most computer-based fluency applications restrict the rate at which students can respond. Computer programs often require subjects to pause following a response to wait for the next practice trial to be presented, thereby suppressing response rate (Binder, 1993; McDade & Goggans, 1993). To avoid this problem in the present experiments, the computer had a fast microprocessor and a program that maximizes system-response time. Thus, progression through trials was virtually uninterrupted. In addition, Binder (1996) suggests that “most computer-based fluency programs suffer from ceilings imposed by component typing skills dysfluencies among most learner populations. High speed voice-recognition technology may offer hope for overcoming this problem in the future (p. 191)”’. In accord with Binder’s suggestion, in the present experiments a vocal response measured by sophisticated voice-recognition technology was used to study fluency.
GENERAL METHOD

Subjects and Setting

Forty-two West Virginia University undergraduate students were recruited via posted advertisement (see Appendix A). Eighteen students served as subjects in Experiment 1; 24 served in Experiment 2. Sixteen males and 26 females between the ages of 17 and 25 years completed the experiments. Prospective subjects were given a brief pretest (shown below) on the binary number system. Only students who were not computer science majors or minors and scored below 25% on the pretest were allowed to participate. One student did not qualify for participation because of a high score on the pretest. Subjects were randomly assigned to experimental conditions (described in detail in the General Method) after the pretest. The following items are the pretest instructions and questions:

Instructions. Prior to being approved for participation in the Fluency with Number Systems on a Computer experiment, you must complete a pretest. For the pretest, you will be asked to complete a short sequence of fill-in-the-blank questions in a paper-and-pencil format. Please answer the questions as best as you can. If you do not know the answer to any of the questions, it is appropriate to guess the answer. When you are finished with the pretest, please notify the experimenter.

Please do not talk during the pretest. Your pretest scores will be used to determine your eligibility for the experiment, but will not be reported to you following completion of the pretest for experimental reasons. No money is available for the pretest. If you have any questions, please ask the experimenter now.

Pretest. For questions 1-12 please write your answer in the space provided.

1. The binary number 00011 is equal to the decimal number _______________.
2. The binary number 00001 is equal to the decimal number _______________.
3. The binary number 10000 is equal to the decimal number _______________.
4. The binary number 01000 is equal to the decimal number _______________.
5. The binary number 00101 is equal to the decimal number _______________.
6. The binary number 01011 is equal to the decimal number _______________.
7. The binary number 10001 is equal to the decimal number _______________.
8. The binary number 11000 is equal to the decimal number _______________.
9. The binary number 01001 is equal to the decimal number _______________.
10. The binary number 00111 is equal to the decimal number _______________.
11. The trinary number 121 is equal to the decimal number _______________.
12. The octal number 24 is equal to the decimal number ________________.

For question 13, you may use the remaining space on this page for your answer.
13. How do you calculate the decimal equivalent of a number from the binary number system?

Experiments were conducted in a 3-m by 6-m windowless room in the Department of Psychology Annex at West Virginia University. No classes were conducted in this building during the experiment, and the location was free from distracting noise. Subjects wore a lightweight microphone headset while they sat at a desk that held an IBM PC-compatible computer, monitor, and keyboard. There was no other furniture in the room.

Apparatus

Experiments were conducted with a Gateway 2000® desktop personal computer with a 33 MHz DX-486 processor, 24 MB of RAM, a 212-MB hard drive, and a 14 in. SVGA color monitor. Experimental programs were written in Turbo Pascal (7.0) and Turbo C (1.5) for MS-DOS. A microphone with noise-canceling and directional-audio properties (Shure Microphones model SM-10A) was attached to the computer’s motherboard by an IBM Multimedia Audio Capture and Playback Adapter board (model M-ACPA) for voice recognition. During experimental sessions in which voice recognition was used, the computer keyboard and mouse were inoperative.

Voice recognition. The voice-recognition software used was the Dragon Systems’ VoiceTools Developers Tool Kit (1.1). This software can be programmed so that function calls can be made to a memory-resident voice-recognition driver to process speech input. When subjects emitted a vocal utterance 25 dB above ambient noise levels, the computer system received an analog signal from the microphone, converted it to a digital format in the M-ACPA audio board, and sent a digital representation of the vocal utterance to the speech-processing
driver. The speech driver compared the digitized pattern with patterns sampled from the subject, then sent an array index of the word spoken (if recognized) to the experimental program.

Digitized patterns were sent to the speech driver during an utterance so that processing and recognition could occur without waiting for the end of an utterance. This procedure allowed voice input to be obtained quickly without slowing the experimental program. For example, during pilot testing, 60 digits were recognized perfectly in a 1-min period. The VoiceTools software also permitted measurement of vocal utterance duration (±20 ms) and amplitude (±1 dB), and provided the speech drivers’ confidence level that a word spoken matched a digitized pattern of that word. In the experiments, only utterances with confidence levels greater than 60% were used. This level was recommended in the VoiceTools software development manual, and was confirmed during pilot testing.

Voice-recognition calibration. Before experiments began, each subject said the words “zero” through “thirty-one” into the microphone, and the digitized patterns of those words were stored for later use in recognition. The standard procedure for training the Dragon Systems speech driver to recognize a particular individual’s speech is to repeat each word in the vocabulary three times (Bamberg & Mandel, 1991). For experimental purposes, the speech driver’s confidence level also was used to calibrate the voice-recognition equipment. Subjects repeated each of the 32 words until the speech driver’s confidence level was above 90% for three consecutive utterances. This calibration procedure took less than 5 min to complete, and ensured the accuracy of the voice-recognition apparatus. In addition, at the beginning of each session, subjects repeated each word until the speech driver’s confidence level for that word was above 90%. Calibration was checked regularly to ensure that changes in utterance topography did not
degrade voice-recognition accuracy. The mean confidence level of voice recognition during the experiments was 96.9% ($N = 96,488; SD = 0.45$).

**General Procedure**

Subjects received the following experimental conditions: Pretest, calibration, accuracy training, fluency building, probe test, and posttest. Pretest, calibration, accuracy training, and posttest conditions were identical for all subjects; independent variables were manipulated only during the fluency-building condition. Pretest and calibration conditions were described above. The other conditions are described in the following sections.

**Accuracy training.** After calibration, subjects read these instructions:

In the following computer-based training session, you will learn how to read numbers from the binary number system. You will interact with the training program by typing on the computer keyboard.

You will be asked a series of questions in the following format. The training topic name, question number (for example, “Question 1 of 37”), and other headings will be shown at the top of the screen. A few lines of text will be shown in the center of the screen, and typically, there are one or more missing words or numbers, which you must supply. When you finish typing your answer to each question, press the “ENTER” key to register your answer. Directly below your answer on the screen, the correct answer will be displayed. You should compare your answer with the correct answer and, if your answer is correct, you should press the “C” key; if it is incorrect, press “I.” (If you are unsure whether your answer is correct or incorrect, please ask the experimenter to help you make that judgment before going on to the next question.) When you have scored your answer, the next question will be presented. Please notify the experimenter when you have completed the training.

After the training session, you will be asked to complete a short sequence of fill-in-the-blank questions in a paper-and-pencil format to assess your understanding of the training material. Please answer the questions as best as you can. If you do not know the answer to any of the questions, it is all right to guess the answer. When you are finished, please notify the experimenter. If you complete less than 80% of these questions correctly, you will need to repeat the training session, so try your best. Your answers to the fill-in-the-blank questions will be scored, but you will not be given your results for experimental reasons.

No money is available for the accuracy training session. If you have any questions, please ask the experimenter now.
If subjects had questions about these instructions, relevant sections were repeated.

Subjects then received a brief sequence of computerized programmed-instruction frames (fill-in-the-blank questions) designed to train subjects to recognize numbers from the binary number system. The computerized accuracy-training program had a similar structure to the program described in Crosbie and Kelly (1994). Subjects completed the same set of instructional materials until they obtained a score greater than 80% (cf. Howell & Lorson-Howell, 1990) on a posttest composed of questions similar to the following items:

1. The binary number 11 is equal to the decimal number __________________.
2. The binary number 1 is equal to the decimal number __________________.
3. The binary number 10000 is equal to the decimal number __________________.
4. The binary number 1000 is equal to the decimal number __________________.
5. The binary number 101 is equal to the decimal number _________________.

An earlier version of the programmed instruction shown in Appendix B was tested on seven graduate students in the Department of Psychology. That program was developed with minimal prompts and large steps. By starting with such a lean program, it was possible to empirically determine the minimum assistance required to produce errorless learning, and then add only that amount of assistance (Markle, 1967). That is, if any student responded incorrectly to a question in the sequence, additional questions were added to the sequence before the question that was answered incorrectly, then the new sequence was presented to another student. This process continued until no questions were answered incorrectly.

Although graduate students were able to answer all questions correctly, subjects were not able to perform as well. Mean percent correct during computerized accuracy training was 83%. Seven subjects repeated accuracy training because they scored below 80% (the threshold) on the posttest. On the second attempt, however, all of those subjects scored above 80% on a parallel form of the posttest.
Fluency building. Following accuracy training, subjects read these instructions:

This experiment involves a series of one-minute intervals in which you will be reading the decimal equivalent of 5-digit binary numbers that will be presented on the computer screen. For example, if you see the binary number 00001, you should say the word one. For each number you read correctly, you will hear a tone and the next number to be read will be presented. If you do not read a number correctly, you will not hear a tone, but the next number will be presented. Sometimes, a message will be displayed on the computer screen after you read a number correctly.

On occasion, you will earn points for reading a number correctly. Each point that you earn is worth 5¢. The computer will keep track of the points you earn, and will display them in a box in the upper-right corner of the computer screen.

Between the one-minute intervals, you will receive feedback about your performance. The computer will display how many numbers you read correctly and incorrectly, as well as the percentage of numbers that you read correctly. In addition, the computer will display a graph of your performance. The number of correct responses you make during each one-minute interval will be displayed in green on the graph; incorrect responses will be shown in red. A blue-colored line also will be displayed on the graph. In general, you will probably earn the highest number of points in this experiment if the green dots (corrects) are near this line.

The computer will inform you when you have completed the series of one-minute intervals. After you finish, you will complete a brief series of tests, then wait two weeks and return for one more test. There is an opportunity to earn additional money on these tests; this opportunity will be explained immediately prior to the tests. You will be paid upon completion of the experiment. It is very important that you attend every session. If you attend all scheduled sessions, you will receive a bonus of $20. If, however, you miss a session without first informing the experimenter, you will not receive this bonus. If you have any questions about these instructions, please ask the experimenter now.

Subjects then completed twenty 1-min practice timings under various reinforcement contingencies (described below) during each 1-hr session. Subjects initiated each timing by saying the word “go.” A randomly-determined sequence of binary numbers (with values between 0 and 31 decimal) were presented one at a time centered on the computer screen.

Although each sequence was determined randomly, all subjects received the same sequence during a given session. The words “Read the decimal equivalent of the following binary number” were displayed above the binary numbers on the computer screen. Responses were
correct if the number spoken matched the decimal equivalent of the binary number displayed on the screen. For example, if 10011 was presented on the screen, the correct vocal response was “nineteen.”

If a response was not recognized by the voice-recognition software (i.e., recognition confidence was less than 60%, or was not in the 32-word vocabulary), the binary number remained on the screen until a word in the vocabulary was spoken. If a response was correct, a 1000-Hz tone was presented for either 20 ms or 100 ms. The next stimulus (i.e., the next binary number to be read) was presented immediately following tone onset. Because subjects could not respond within 100 ms, feedback tones did not artificially constrain response rate (cf. Binder, 1996) by imposing a delay between successive trials. If a response was incorrect, no tone was presented, and the next trial began immediately. After 60 s, the practice timing stopped regardless of subjects’ behavior. After each timing, the following information for that timing was presented on the computer screen: the number of reinforcers obtained, percent of responses answered correctly, and the number of correct and incorrect responses made. The number of reinforcers obtained across all timings also was displayed.

Between fluency-building timings, a graph of the number of correct and incorrect responses was displayed on the computer screen for 15 s. This graph had a semi-logarithmic vertical axis similar to those used in precision teaching (Lindsley, 1990b), and 1-min timings were displayed along the horizontal axis. Celeration target lines (described below) also were displayed on the graph. The sample graph shown below is a close approximation of graphs presented to subjects.
Figure 1. Sample feedback chart shown to subjects between fluency-building timings.

Reinforcement Contingencies

Subjects received points exchangeable for money during fluency-building timings (5 cents per point) and REAPS posttests (4 cents per point; described below). Exchange rates were arranged such that subjects would earn approximately $6 per hour of participation. Subjects also received a $20 bonus for attending all scheduled sessions. Two subjects (H13 and H53) forfeited this bonus for missing a session without informing the experimenter beforehand. Subjects were paid at the end of their participation in the experiment.

A point counter was displayed in the upper-right corner of the computer screen during fluency building, but not during REAPS posttests. Increments of the counter were correlated with an auditory stimulus (described below).

Reinforcement schedule. Responding was reinforced on a variable interval (VI) 10 s schedule, where the only responses eligible for reinforcement were those with interresponse
times (IRTs) within a programmed range. On a VI schedule of reinforcement, reinforcers become available after intervals of varying lengths (around an average) have elapsed. When the interval has elapsed, the next response is followed by the reinforcer. The 10-s interval used in the present experiments allowed subjects the opportunity to receive an average of five points (worth 25 cents) during a 1-min timing. Fleshler and Hoffman’s (1962) method was used to produce intervals.

When the VI interval had elapsed, the next correct response with an IRT within a programmed range of IRTs was reinforced. The limits of the programmed range of IRTs eligible for reinforcement were 25% above and 25% below the target IRT for any given timing. For example, if the target IRT for a particular timing was 4 s, the programmed range of IRTs eligible for reinforcement was 3 s – 5 s. Given this example, a reinforcer would be presented after completion of a correct response that followed the end of the previous correct response by 3 s to 5 s. IRTs were measured from the end of a correct utterance to the end of the next correct utterance (i.e., response latency plus utterance duration).

The parameters of the IRT range varied across experimental sessions according to celeration targets described below. The purpose of this procedure was to allow response rates, and therefore celeration rates, to be controlled. The purpose of the VI component of the reinforcement schedule was to ensure that reinforcement density would not be confounded with experimental conditions.

**Celeration targets.** Reinforcement was arranged such that response rates tracked a specified celeration target (x2 or x1.25). The range of IRTs eligible for reinforcement was determined using the following procedure. The target response rate for the initial 1-min timing during the fluency-building phase was 10 responses per min. The target IRT for this response
rate is 6 s, and the programmed range of IRTs eligible for reinforcement is 4.5 s - 7.5 s. As the
target rate increased across sessions, the range of IRTs eligible for reinforcement was adjusted
such that response rate matched the desired celeration. For example, the desired response rate
after 20 timings for a x2 celeration target is 20 responses per min. Given a target rate of 20
responses per min, the target IRT is 3 s, and the range of IRTs eligible for reinforcement is
2.25 s - 3.75 s (i.e., 3 s ± 25%).

Table 1

<table>
<thead>
<tr>
<th>Session</th>
<th>X1.25</th>
<th>X2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Target</td>
</tr>
<tr>
<td>1</td>
<td>4.50</td>
<td>6.00</td>
</tr>
<tr>
<td>6</td>
<td>4.34</td>
<td>5.79</td>
</tr>
<tr>
<td>11</td>
<td>4.18</td>
<td>5.57</td>
</tr>
<tr>
<td>16</td>
<td>4.03</td>
<td>5.37</td>
</tr>
<tr>
<td>21</td>
<td>3.88</td>
<td>5.17</td>
</tr>
<tr>
<td>26</td>
<td>3.74</td>
<td>4.99</td>
</tr>
<tr>
<td>31</td>
<td>3.60</td>
<td>4.80</td>
</tr>
</tbody>
</table>

Table 1 shows various target IRTs and corresponding ranges of IRTs eligible for
reinforcement for both celeration targets. Each celeration target in Table 1 shows the lowest
IRT, the target IRT, and the highest IRT eligible for reinforcement for every fifth session from
Timings 1 through 31. Note that the target IRT for Session 21 for the x2 celeration target (3 s) is
half the target IRT for the first session (6 s). A halving of the target IRT corresponds to a
doubling in the target response rate, which allows the celeration target of x2 to be maintained.
## Table 2

### Summary of Tonal Frequency and Duration, Textual Feedback Messages, and Reinforcement Conditions

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th></th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within IRT range</td>
<td>Outside IRT range</td>
<td>IRT &gt; range</td>
</tr>
<tr>
<td>VI elapsed</td>
<td>Yes</td>
<td>No</td>
<td>IRT &lt; range</td>
</tr>
<tr>
<td>Tone frequency (Hz)</td>
<td>1000</td>
<td>1000</td>
<td>200</td>
</tr>
<tr>
<td>Tone duration (ms)</td>
<td>100</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Text message</td>
<td>“Point earned”</td>
<td>--</td>
<td>“Too slow”</td>
</tr>
<tr>
<td>Point earned?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Feedback tones and messages.** Table 2 shows a summary of feedback tones and messages presented to subjects. Correct responses that were made after the VI interval had elapsed and had an IRT within the reinforceable range were followed by a 1000-Hz tone for 100 ms, “point earned” displayed on the screen one line below the binary number, and an increase of the point counter. Correct responses with IRTs within the reinforceable range, but that were not eligible for reinforcement according to the VI schedule, were followed by a 1000-Hz tone for 20 ms without any text message or increase in points. The frequencies for both tones correlated with correct responses were identical. Tone duration was different so that only correct responses that also met the VI schedule requirements would be correlated with a 100-ms tone and increment of the point counter.

Correct responses with IRTs greater than the upper limit of the range of IRTs eligible for reinforcement (i.e., IRTs that were too long) were followed by a 200-Hz tone for 20 ms and the message “Too slow.” Correct responses with IRTs less than the lower limit of the range of IRTs would be followed by a 1000-Hz tone for 100 ms without any text message or increase in points.
eligible for reinforcement (i.e., IRTs that were too short) were followed by an 1800-Hz tone for 20 ms and the message “Too fast.” Frequencies of these tones were 800 Hz above and below the frequency of tones produced by correct responses with reinforceable IRTs. There were no programmed consequences for incorrect responses.

Tests for REAPS Outcomes

Probe tests. The purpose of probe tests was to assess whether subjects could combine the ability to recognize numbers from the binary number system with the ability to solve mathematical problems (application), and whether distracting stimuli slows responding (stability). Subjects in Experiment 1 received both probe tests after every five fluency-building timings. By measuring each subject’s performance in this manner, it could be shown whether different rates during fluency building were correlated with performance on application and stability probes. Probe tests were administered during Experiment 1 only, because data from Experiment 1 suggested that testing these fluency outcomes with probe tests was not useful (see Results and Discussion section in Experiment 1 for details). Other REAPS tests were not assessed until after the fluency-building phase because extra practice on the experimental task might have interfered with performance on fluency-building timings.

Instructions were read to subjects before the first probe tests, then were posted near the computer for future reference. Subjects received these instructions before the first application probe (note that Paragraph 2 was omitted during probe tests, but was used during posttests):

This test involves answering some addition, subtraction, and multiplication math problems using both binary and decimal numbers presented on the computer screen. Each math problem is different, so make sure you read the instructions on the screen for each question. Type your answer to each question, then press enter to advance to the next question. You will not be given any feedback as to whether your answers are correct or incorrect. The computer will not make any beeps or display any messages that will tell you if you answered correctly, but the computer will keep track of your responses.
You will earn $0.04 for every correct response. That does not mean you will lose $0.04 if you answer a question incorrectly, it just means that you will lose the opportunity to earn the $0.04 that you would have received, had you answered the question correctly. If you have any questions about these instructions, please ask the experimenter now.

Questions from the application tests were presented on the computer in a fill-in-the-blank format similar to that used in the accuracy-training phase. The computer program used for the application tests was identical to the program used for accuracy training, with the exceptions that (a) subjects were not allowed to see the correct answer, and (b) subjects were not given feedback concerning the accuracy of their responses. Addition, subtraction, and multiplication math problems with combinations of both binary and decimal numbers were used. The following items are a sample of questions asked during application probe tests:

1. Fill in the blank with a Binary number: 11 + 8 = _______
2. Fill in the blank with a Binary number: 117 - 92 = _______
3. Fill in the blank with a Binary number: 01000 + 10110 = _______
4. Fill in the blank with a Decimal number: 11001 + 10111 = _______
5. Fill in the blank with a Binary number: 00011 * 00101 = _______
6. Fill in the blank with a Decimal number: 11110 * 00010 = _______

Subjects received these instructions before the first stability probe (note that Paragraph 2 was omitted during probe tests but was used during posttests):

This test involves reading the decimal equivalent of binary numbers presented on the computer screen. It is very similar to what you have done in the past. Here is what is different for this test: First, you will not be given any feedback as to whether your answers are correct or incorrect. The computer will not make any beeps or display any messages that will tell you if you read a number correctly, but the computer will keep track of your responses. Second, there will be more than one number displayed on the screen. It will be obvious which number you should read, because it will be a different color than the others and it will be presented in the center of the screen.

In addition, you will earn $0.04 for every correct response. That does not mean you will lose $0.04 if you read a number incorrectly, it just means that you will lose the opportunity to earn the $0.04 that you would have received, had you read the number correctly. Because you are earning $0.04 for each correct response and this is a timed test, it is in your best interest to go as fast as you can to earn the most money.

If you have any questions about these instructions, please ask the experimenter now.
The probe test for performance stability involved exposing subjects to distraction conditions to assess how much response rate and accuracy would decrease. Fluent performance should be affected only minimally by such distraction (Binder, 1993). Subjects completed a 1-min timing similar to those used during fluency building with two differences. Instead of presenting only one stimulus centered on the computer screen, on each probe trial the screen was filled with approximately 100 random decimal numbers between 0 and 31 displayed in light gray. These distraction stimuli were the same size as the primary stimulus (i.e., binary number), and had one blank space between them. The binary number to be read by subjects was displayed in white in the center of the screen. In addition, subjects were not given feedback about the accuracy of their responses so that the only training was during fluency building.

Posttests. Following the fluency-building condition, subjects completed 4 of 5 posttests; a fifth test (retention test) was given two weeks after the other posttests. The VI schedule and IRT range were not used in any posttest. Subjects earned 4 cents for each correct response on all posttests.

Retention in precision teaching is typically assessed over a period of two weeks without practicing a skill (Johnson & Layng, 1994); that was the interval used in the present experiments. Application and performance stability posttests were identical to probe tests except that subjects earned money for their performance. Endurance and adduction posttests also were administered.

The endurance posttest was identical to fluency-building timings except that the duration was 3-min, no feedback was given to subjects about the accuracy of their responses, and each correct answer was worth 4 cents. The purpose of the endurance test was to assess whether response rate or accuracy would decrease if subjects were required to read binary numbers for a
longer time than was used during fluency building. The following instructions for the endurance test were read to subjects:

This test involves reading the decimal equivalent of binary numbers presented on the computer screen. It is very similar to what you have done in the past. Here is what is different for this test: First, you will not be given any feedback as to whether your answers are correct or incorrect. The computer will not make any beeps or display any messages that will tell you if you read a number correctly, but the computer will keep track of your responses. Second, you will be reading numbers for 3 consecutive minutes instead of 1 minute. In addition, you will earn $0.04 for every correct response. That does not mean you will lose $0.04 if you read a number incorrectly, it just means that you will lose the opportunity to earn the $0.04 that you would have received, had you read the number correctly.

Because you are earning $0.04 for each correct response and this is a timed test, it is in your best interest to go as fast as you can to earn the most money. If you have any questions about these instructions, please ask the experimenter now.

The adduction posttest required subjects to convert numbers from the trinary (Base 3) and octal (Base 8) number systems to decimal numbers. The purpose of the adduction posttest was to assess the extent to which subjects could read numbers from a set of untrained number systems after learning binary. Here are the adduction posttest instructions and questions:

In this experiment, you learned how to read numbers from the binary number system (Base 2). To read a number from any number system, consider each digit in the number, the column location of that digit, and the base of the number system.

Please write the decimal equivalent of the following trinary numbers (Base 3) in the space provided.

- 210 Trinary = _____________ decimal
- 111 Trinary = _____________ decimal
- 022 Trinary = _____________ decimal

Please write the decimal equivalent of the following octal numbers (Base 8) in the space provided.

- 27 Octal = _____________ decimal
- 17 Octal = _____________ decimal
- 30 Octal = _____________ decimal
To assess retention, subjects completed a 1-min timing that was identical to fluency-building timings except that no feedback was given to subjects about the accuracy of their responses. The purpose of the retention posttest was to assess whether response rate or accuracy would decrease relative to terminal rates obtained during the fluency-building phase if subjects were not given an opportunity to practice reading binary numbers using the experimental apparatus for an extended period (two weeks). Here are the instructions for the retention test:

This test involves reading the decimal equivalent of binary numbers presented on the computer screen. It is very similar to what you have done in the past. Here is what is different for this test: First, you will not be given any feedback as to whether your answers are correct or incorrect. The computer will not make any beeps or display any messages that will tell you if you read a number correctly, but the computer will keep track of your responses. In addition, you will earn $0.04 for every correct response. That does not mean you will lose $0.04 if you read a number incorrectly, it just means that you will lose the opportunity to earn the $0.04 that you would have received, had you read the number correctly.

Because you are earning $0.04 for each correct response and this is a timed test, it is in your best interest to go as fast as you can to earn the most money. If you have any questions about these instructions, please ask the experimenter now.

Debriefing. Following completion of the posttests, subjects were given a debriefing questionnaire. Subjects were told that the purpose of the experiment was to assess how different types of feedback affect learning, and that details of the experiment could be obtained from the experimenter once all subjects had completed the experiment. No subjects contacted the experimenter to ask for this information. Here is the debriefing questionnaire:

1. What was your strategy for earning the most points during this experiment?
2. What was the purpose of showing you graphs between each one-minute period in which you were reading the decimal equivalent of binary numbers?
3. What was the purpose of the tests at the end of the experiment?
4. What did you like about this experiment?
5. What didn’t you like about this experiment, or what would you do differently?
6. Is there anything else that you would like to comment on about your participation in this experiment?
EXPERIMENT 1: Fluency Aims

Method

Procedure

The purpose of the first experiment was to determine the lowest terminal rate of reading binary numbers at which subjects demonstrate fluency according to performance on REAPS posttests. That target rate was used as the criterion for termination of the fluency-building phase in Experiment 2.

The target rate was determined by conducting a parametric analysis of the effects of terminal response rates on performance on REAPS posttests. After completing pretest and accuracy-training phases, subjects repeated fluency-building timings until they met one of four rate criteria (25, 30, 35, or 40 responses per minute), then REAPS posttests were administered. Responding was reinforced using a x2 celeration target to determine the optimal rate under ideal conditions. Consequently, terminal rate was the independent variable, and performance on REAPS posttests was the dependent variable.

Performance on the retention, endurance, and performance stability posttests was operationally defined as satisfactory if (a) the mean response rate per minute during each of these tests was at least 85% of the mean rate of the last 5 fluency-building timings, and (b) accuracy on each test was above 90% correct. Performance on the application posttest was operationally defined as satisfactory if accuracy on the test was above 85% correct. The adduction posttest also was administered but was not used to determine fluency, because the conceptual adequacy of adduction has not been assessed.
Results and Discussion

The optimal fluency aim (i.e., rate criterion) for learning binary numbers with the present procedures was chosen from the set of 25, 30, 35, and 40 responses per minute based on two criteria: (a) the effectiveness of each terminal rate in predicting criterion performance on REAPS outcomes, and (b) the efficiency of performance within each group during fluency building. Effectiveness in predicting REAPS outcomes was the primary criterion for selecting the fluency aim for Experiment 2. Hence, data on effectiveness are discussed first. Following selection of the optimal fluency aim, application and stability performance are discussed for subjects who achieved that fluency aim.

Effectiveness

Table 3 shows a summary of performance on REAPS posttests for subjects in each fluency-aim group. Mean correct responses per minute from the last five timings during fluency building is shown in Column 3 so that comparisons can be made with responses per minute on retention, endurance, and stability posttests. Correct responses per minute plus the percentage of the mean of the last five fluency-building timings is shown for retention, endurance, and stability tests. Percent correct also is shown in Table 3 for application and adduction posttests, although performance on the adduction posttest was not used to judge effectiveness.

The data shown in Table 3 were used to judge the effectiveness of each terminal rate. The percent of the last five timings during fluency building was used to judge effectiveness for retention, endurance, and stability posttests. Scores below 85% on the percent of the last five timings (i.e., mean correct rate on posttests divided by mean correct rate from the last five fluency-building timings) did not satisfy the criterion for fluency.
Table 3

Experiment 1: Summary of REAPS Posttest Data.

<table>
<thead>
<tr>
<th>Rate/Min</th>
<th>Subject</th>
<th>Last 5</th>
<th>Retention</th>
<th>% Last 5</th>
<th>Endurance</th>
<th>% Last 5</th>
<th>Stability</th>
<th>% Last 5</th>
<th>Application</th>
<th>Adduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>H15</td>
<td>22.0</td>
<td>27</td>
<td>123</td>
<td>24.0</td>
<td>109</td>
<td>16</td>
<td>73</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H18</td>
<td>21.0</td>
<td>27</td>
<td>129</td>
<td>20.3</td>
<td>97</td>
<td>20</td>
<td>95</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H19</td>
<td>22.8</td>
<td>28</td>
<td>123</td>
<td>22.3</td>
<td>98</td>
<td>23</td>
<td>101</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H20</td>
<td>21.6</td>
<td>30</td>
<td>139</td>
<td>22.0</td>
<td>102</td>
<td>18</td>
<td>83</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>H22</td>
<td>23.6</td>
<td>26</td>
<td>110</td>
<td>23.0</td>
<td>97</td>
<td>16</td>
<td>68</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>H07</td>
<td>26.2</td>
<td>30</td>
<td>115</td>
<td>26.3</td>
<td>100</td>
<td>23</td>
<td>88</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>H08</td>
<td>25.4</td>
<td>32</td>
<td>126</td>
<td>24.0</td>
<td>94</td>
<td>25</td>
<td>98</td>
<td>88</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>H09</td>
<td>26.4</td>
<td>30</td>
<td>114</td>
<td>26.0</td>
<td>98</td>
<td>26</td>
<td>98</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H11</td>
<td>28.4</td>
<td>32</td>
<td>113</td>
<td>29.0</td>
<td>102</td>
<td>26</td>
<td>92</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>H12</td>
<td>27.4</td>
<td>32</td>
<td>117</td>
<td>26.6</td>
<td>97</td>
<td>24</td>
<td>88</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td>35</td>
<td>H02</td>
<td>31.2</td>
<td>33</td>
<td>106</td>
<td>31.0</td>
<td>99</td>
<td>32</td>
<td>103</td>
<td>88</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>H03</td>
<td>23.8</td>
<td>28</td>
<td>118</td>
<td>22.3</td>
<td>94</td>
<td>21</td>
<td>88</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H04</td>
<td>30.2</td>
<td>27</td>
<td>90</td>
<td>29.6</td>
<td>98</td>
<td>26</td>
<td>86</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>H05</td>
<td>31.8</td>
<td>34</td>
<td>107</td>
<td>30.3</td>
<td>95</td>
<td>29</td>
<td>91</td>
<td>88</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>H06</td>
<td>31.4</td>
<td>34</td>
<td>108</td>
<td>28.3</td>
<td>90</td>
<td>30</td>
<td>96</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>H13</td>
<td>30.2</td>
<td>37</td>
<td>123</td>
<td>31.3</td>
<td>104</td>
<td>26</td>
<td>86</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H14</td>
<td>35.4</td>
<td>38</td>
<td>107</td>
<td>35.0</td>
<td>99</td>
<td>31</td>
<td>88</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H16</td>
<td>34.2</td>
<td>37</td>
<td>108</td>
<td>34.3</td>
<td>100</td>
<td>30</td>
<td>88</td>
<td>100</td>
<td>67</td>
</tr>
</tbody>
</table>

Note. Last 5 is the mean rate of correct responses from the last five fluency-building timings. Retention is the number of correct responses per minute on the retention posttest. Endurance is mean correct responses per minute on the 3-min endurance posttest. Stability is the number of correct responses per minute on the stability posttest. Application is the percent correct on the application posttest. Adduction is the percent correct on the adduction posttest.
Scores on the retention posttest were higher than the rate criterion for subjects in the 25 and 30 responses per minute group, but were lower than the rate criterion for subjects in the 35 and 40 responses per minute group. Hence, although all subjects met the criterion for fluency on the retention posttest (85%), subjects in the 25 and 30 responses per minute group had superior performance relative to their terminal rate condition. With the exception of one subject in the 35 responses per minute group (H04), retention scores were higher than the mean of the last five fluency-building timings (percentage last 5 = 90).

All subjects met the criteria for fluent performance on the Endurance test (85%). Each subject except H04 had a better score on the retention test than on the endurance test.

The only subjects with sub-criterion performance on the stability posttest (<85%) were H15, H20, and H22 (see shaded bold text in Table 3). Each of these subjects was in the 25 responses per minute group. Subject H15 made 16 correct responses on the stability posttest, which was 73% of the mean response rate of the last 5 fluency-building timings. Subject H20 made 18 correct responses on the stability posttest, which was 83% of the mean response rate of the last 5 fluency-building timings. Subject H22 made 16 correct responses on the stability posttest, which was 68% of the mean response rate of the last 5 fluency-building timings.

The only subjects with sub-criterion performance on the application posttest (<85%) were H15 and H22 (see shaded bold text in Table 3). Each of these subjects was in the 25 responses per minute group. For these subjects, 38% and 75% of questions on the application posttest were answered correctly (H15 and H22, respectively). The lowest score on the application posttest for all other subjects was 88%.

**Summary of REAPS performance.** All subjects in Experiment 1 met the criteria for acceptable performance on retention and endurance tests. In addition, all subjects in the 30, 35,
and 40 responses per minute groups met the criteria for acceptable performance on stability and application tests. Each of these terminal rates therefore produced fluent performance on the experimental task according to the present criteria. Because stability and application performance from some subjects in the 25 responses per minute group was not acceptable, that terminal rate was not considered for Experiment 2; all other rates were acceptable.

**Efficiency**

Figures 2 through 5 show response rate during the fluency-building condition for the groups with rate criteria of 25, 30, 35, and 40 responses per minute, respectively. These figures are similar to the feedback graphs that were shown to subjects between timings. Circles and triangles show correct and incorrect responses, respectively. These figures also show the fluency aim (dashed horizontal line) and the ideal response rate that could be produced by the celeration contingency (dotted line). For these data, maximum efficiency would be demonstrated if correct responses fell directly on the ideal response-rate line with few incorrect responses.

Figure 2 shows data from the 25 responses per minute group. H15’s correct response rate was far from the ideal rate throughout the experiment. H18 and H19’s correct response rate was close to the ideal rate in the first timings, but they each had difficulty reaching the terminal rate. H18 also had more incorrect responses toward the end of the experiment. H20 and H22 had efficient performance throughout fluency building.

Figures 3 and 4 show data from the 30 and 35 responses per minute groups, respectively. Visual analysis of correct responses shows that performance of subjects in the 35 responses per minute group (especially H02, H04, and H05) generally was less efficient than that of subjects in the 30 responses per minute group. That is, rate of correct responses for subjects in the 35 responses per minute group did not track the ideal rate line as well as did correct response rate
Figure 2. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 25 responses per minute rate criterion group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic.
Figure 3. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 30 responses per minute rate criterion group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic.
Figure 4. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 35 responses per minute rate criterion group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic.
Figure 5. Experiment 1: Rate of correct and incorrect responses during fluency building for subjects in the 40 responses per minute rate criterion group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic. Arrows in the top panel show changes in the celeration contingency.
for subjects in the 30 responses per minute group, and it took longer to reach the terminal rate. In general, however, performance of subjects in both of these groups was efficient.

Figure 5 shows that efficiency in the 40 responses per minute group is low. In fact, the celeration contingency was modified for the first subject who completed the experiment in that group (H13), because it did not appear that his performance would be able to track the ideal rate. The range of IRTs eligible for reinforcement for H13 was modified following Timing 66, and again following Timing 129 (see arrows in the top panel of Figure 5). The first modification involved increasing the number of timings in which response rate doubles from 20 to 30; the second increased that number to 40. Neither modification increased H13’s efficiency. H14 and H16 received the first modified celeration contingency (i.e., 30 timings required for a doubling in rate to occur). H14’s performance was efficient for approximately 50 timings, but did not reach 40 correct responses per minute until Timing 115. Rate of correct responses for H16 did not reach 40 per minute until Timing 198, and H16 made more incorrect responses toward the end of the experiment. Because performance at 40 responses per minute was so inefficient, this rate was not considered further for Experiment 2.

The following analyses also were performed to compare the efficiency of the 30 and 35 responses per minute groups: the percent of (a) correct responses that had IRTs within the programmed IRT range, (b) reinforcers earned per opportunity, and (c) timings before subjects reached 0% of responses with IRTs within the programmed range during fluency building.

Figures 6 and 7 show the percent of responses during fluency building with IRTs within the programmed range for the 30 and 35 responses per minute groups, respectively. Higher percentages indicate higher efficiency. The inverted U-shaped pattern shown in these figures suggests an initial “warm-up” period (i.e., a period where responding gradually came under
control of the experimental contingencies), followed by a period of efficient responding, and ending with a gradual decline in efficiency toward the end of the experiment.

Figure 7 shows that three subjects from the 35 responses per minute group (H02, H04, and H05) had extended periods when no responses had IRTs within the reinforceable range, whereas only one subject in the 30 responses per minute group had that pattern (H11; see Figure 6). To assess the effect this had on the effectiveness of the reinforcement contingency, an analysis of reinforcers per opportunity was conducted.

Figures 8 and 9 show the results of the analysis of reinforcers earned per opportunity for the 30 and 35 responses per minute groups, respectively. Subjects could earn up to five points in each fluency-building timing. Scores of 100% in Figures 8 and 9 indicate that all available reinforcers were earned during that timing. The inverted U-shape pattern shown in Figures 8 and 9 is similar to that shown in Figures 6 and 7. Hence, as subjects made fewer responses with IRTs in the reinforceable range, the probability of receiving reinforcers also decreased. In addition, all subjects in the 35 responses per minute group (Figure 9) had extended periods when no reinforcers were earned.

The analysis of the percentage of responses within the reinforceable IRT range plus the analysis of reinforcers earned per opportunity suggests that the 30 responses per minute fluency aim produced more efficient responding than did the 35 responses per minute fluency aim. Responding of subjects in the 30 responses per minute group satisfied the requirements of the reinforcement contingency more frequently than did responding of subjects in the 35 responses per minute group. That is, subjects in the 30 responses per minute group spent more timings earning reinforcement, because they also had more timings with IRTs within the reinforceable range.
Figure 6. Experiment 1: Percent of responses during 1-min timings with IRTs within the range of IRTs eligible for reinforcement for the 30 responses per minute group.
Figure 7. Experiment 1: Percent of responses during 1-min timings with IRTs within the range of IRTs eligible for reinforcement for the 35 responses per minute group.
Figure 8. Experiment 1: Percentage of reinforcers earned per available opportunity for subjects in the 30 responses per minute group.
Figure 9. Experiment 1: Percentage of reinforcers earned per available opportunity for subjects in the 35 responses per minute group.
The relative proportion of the experiment before subjects stopped earning reinforcement because IRTs were not within the reinforceable range is a good indicator of the efficiency of the experimental conditions. The percentage of the experiment before subjects in the 30 responses per minute group had no responses within the IRT range was 98%, 97%, and 75% (for H07, H09, and H11 respectively; H08 and H12 never reached this level), whereas for subjects in the 35 responses per minute group this percentage was 52%, 82%, 72%, 56%, and 91% (for H02, H03, H04, H05, and H06, respectively). Subjects in the 30 responses per minute group therefore spent more of the experiment earning reinforcement than did subjects in the 35 responses per minute group.

**Summary of effectiveness and efficiency.** In the search for the optimal fluency aim, 25 responses per minute was discarded because some subjects in that group performed poorly on stability and application posttests. Forty responses per minute was discarded because subjects were unable to maintain efficient responding during fluency building despite modifications in the celeration contingency. Both 30 and 35 responses per minute produced effective performance on REAPS posttests, but 30 responses per minute was more efficient than was 35 responses per minute. Consequently, 30 responses per minute was the optimal rate, and therefore was used in Experiment 2.

**Probe performance.** Figure 10 shows performance on probe tests during fluency building for subjects in the 30 responses per minute group. Left panels show response rate on stability probes (unfilled circles) and mean response rate for blocks of five timings from fluency building (filled circles). Panels on the right side of Figure 10 show percent correct on application probes (filled triangles).
In general, rate of correct responses on stability probes tracked rate of correct responses during fluency building (cf. open and closed circles). Visual inspection of stability-probe data suggests that rate of correct responses was higher during stability probes than the mean fluency-building rate for the first five blocks of timings, but was not different thereafter. Presumably, response rate under distraction conditions should be less than rate under non-distraction conditions, but the non-distraction condition present for fluency building in Experiment 1 also included a celeration contingency. Hence, it is possible that subjects might have gone faster without the celeration contingency and therefore had higher response rates than during distraction during the first few blocks of timings. In addition, it also is possible that the celeration contingency generalized to stability probes, thereby influencing response rate. Regardless, it did not appear that performance during stability tests was affected in any systematic way by fluency.

Application-probe performance from most subjects (i.e., all but H07) was 7 out of 8 correct responses by the second application probe. Application-probe performance remained high for the remainder of the experiment for all subjects except H11. It is interesting to note that this was the only subject in the 30 responses per minute group who had no IRTs within the programmed range during fluency building, and that her performance on application probes decreased to 50% correct immediately following the block of timings in which 0% was obtained (see Figures 6 and 10). It does not appear from Figure 10 that performance during application-probe tests was affected in any systematic way by fluency. In fact, accuracy levels were high early in the experiment and generally remained high throughout.

Performance on probe tests was not affected in any systematic way by fluency-building procedures, yet the only REAPS posttests that distinguished performance across experimental
groups were application and stability posttests. These fluency outcomes were therefore not tested as probes during Experiment 2, but remained in the REAPS posttests.

Figure 10. Experiment 1: Performance on probe tests for subjects in the 30 responses per minute group. Left panels show rate of correct responses on stability probes (unfilled circles) with mean rate of correct responses from blocks of 5 fluency-building timings. Right panels show percent correct on application probes.
EXPERIMENT 2: Celeration

Method

Procedure

The purpose of Experiment 2 was to determine whether different celeration rates are correlated with better performance on REAPS posttests. The rate criterion that predicted fluent performance on all REAPS posttests in Experiment 1 was used as the terminal rate for all subjects in Experiment 2. Hence, celeration rate was the independent variable in Experiment 2, and performance on REAPS posttests was the dependent variable. A total of 11 subjects were assigned to the x2 condition; 12 subjects were assigned to the x1.25 condition.

Subjects were given a pretest and accuracy training in the same manner as Experiment 1, then completed fluency-building timings under reinforcement contingencies designed to maintain the appropriate celeration targets (x2 and x1.25) until performance reached the rate criterion (30 responses per minute), then completed REAPS posttests. Performance on posttest measures was used to compare performance across groups who attained the same terminal response rate under different celeration targets.

Results and Discussion

Although the efficiency and effectiveness of the current experimental procedures (i.e., the x2 celeration contingency and 30 responses per minute fluency aim) was validated in Experiment 1, it was important to ensure that similar levels of experimental control were obtained in Experiment 2. Hence, effectiveness and efficiency of performance in Experiment 2 are examined below.

x2 celeration rate. Figures 11 and 12 show performance from the fluency-building component for subjects in the x2 celeration group. Rate of correct (filled circles) and incorrect
responses (filled triangles), the 30 responses per minute fluency aim (dashed horizontal line), and ideal rate according to the reinforcement contingency (dotted line) are shown across 1-min timings. These data are similar to corresponding data from subjects in the 30 responses per minute group in Experiment 1 (see Figure 3), with rate of correct responding tracking the ideal rate line for most subjects (except H53; see Figure 12).

Figures 13 and 14 show efficiency data for subjects in the x2 celeration group. The pattern of results shown in these figures also is similar to that shown in Figure 6 from Experiment 1. That is, efficiency of subjects’ responding is low in the first few timings, then increases, and finally declines toward the end of the experiment. In addition, only three subjects out of 11 (H30, H32, and H34) had one or more timings late in the experiment in which no responses had IRTs within the programmed range of reinforceable IRTs. The percentage of the experiment before no IRT was within the reinforceable range was 75%, 79%, and 79% (for H30, H32, and H34, respectively; no other subject reached 0%).

The percentage of the experiment before subjects in the 30 responses per minute group in Experiment 1 had no responses within the programmed range was 98%, 97%, and 75% (for H07, H09, and H11, respectively). Hence, efficiency was similar for subjects in the 30 responses per minute group in Experiment 1 and the x2 celeration group in Experiment 2.

x1.25 celeration rate. Figures 15 and 16 show performance from the fluency-building condition for subjects in the x1.25 celeration group. Rate of correct (filled circles) and incorrect responses (filled triangles), the 30 responses per minute fluency aim (dashed horizontal line), and ideal rate according to the reinforcement contingency (dotted line) are shown across 1-min timings. In general, these figures show that rate of correct responding tracks the ideal rate
Figure 11. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x2 celeration group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic.
Figure 12. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x2 celeration group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x2 celeration. Note that the vertical axis is logarithmic.
Figure 13. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x2 celeration group.
Figure 14. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x2 celeration group.
line, which indicates that responding was controlled by the experimental contingencies in a similar fashion to the x2 group.

Efficiency data are shown for subjects in the x1.25 celeration group in Figures 17 and 18. The familiar inverted U-shaped function is the predominant pattern in these figures, yet not all subjects showed this pattern (especially H31, H33, and H37 in Figure 17, and H57 in Figure 18). Eight of 12 subjects in the x1.25 celeration group (i.e., H35, H38, H47, H48, H51, H52, H55, and H59) had more than 10 timings with no IRTs within the reinforceable range. The percentage of the experiment before IRTs within the reinforceable range reached 0% for these eight subjects was 58%, 24%, 53%, 50%, 43%, 42%, 37%, and 36% (for H35, H38, H47, H48, H51, H52, H55, and H59 respectively; performance of other subjects in this group never reached that level).

Hence, responding during the fluency-building condition was not as efficient for subjects in the x1.25 celeration group as it was for subjects in the x2 celeration group.

**REAPS outcomes.** Table 4 shows a summary of scores on REAPS posttests for all subjects in Experiment 2. Means and standard deviations are shown below each section, and t values with corresponding significance values are displayed at the bottom of each column. For example, the mean retention score for subjects in the x2 celeration group was 25 (SD = 4), and the mean retention score for subjects in the x1.25 celeration group was 26 (SD = 3). The mean score for the x2 celeration group was not significantly greater than that of the x1.25 celeration group (t [20] = -0.56, p = 0.29).

A one-tailed t test was conducted for REAPS posttests instead of a two-tailed test, because the x2 celeration value is considered superior by precision teachers (Binder, 1996). No between-group differences were found in performance on any REAPS posttest (see the last two rows in Table 4 for t and p values). Furthermore, the t values are so low that an infinite number
Figure 15. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x1.25 celeration group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x1.25 celeration. Note that the vertical axis is logarithmic.
Figure 16. Experiment 2: Rate of correct and incorrect responses during fluency building for subjects in the x1.25 celeration group. Circles show rate of correct responses; triangles show rate of incorrect responses. Dashed lines indicate the fluency aim, and dotted lines indicate the target rate for the x1.25 celeration. Note that the vertical axis is logarithmic.
Figure 17. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x1.25 celeration group.
Figure 18. Experiment 2: Percent of responses during a 1-min timing with IRTs within the range of IRTs eligible for reinforcement for the x1.25 celeration group.
Table 4

**Experiment 2: Scores on REAPS Posttests.**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Retention</th>
<th>Endurance</th>
<th>Stability</th>
<th>Application</th>
<th>Adduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Celeration: x2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H28</td>
<td>28</td>
<td>88</td>
<td>25</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>H30</td>
<td>30</td>
<td>89</td>
<td>24</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>H32</td>
<td>15</td>
<td>68</td>
<td>19</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>H34</td>
<td>26</td>
<td>71</td>
<td>25</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>H36</td>
<td>26</td>
<td>75</td>
<td>26</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>H49</td>
<td>23</td>
<td>86</td>
<td>22</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>H50</td>
<td>25</td>
<td>74</td>
<td>27</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>H58</td>
<td>28</td>
<td>87</td>
<td>22</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>H60</td>
<td>27</td>
<td>85</td>
<td>21</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>H53</td>
<td>26</td>
<td>81</td>
<td>28</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>H56</td>
<td>24</td>
<td>61</td>
<td>26</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>M</td>
<td>25</td>
<td>79</td>
<td>24</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>SD</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Celeration: x1.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H31</td>
<td>24</td>
<td>79</td>
<td>25</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>H33</td>
<td>25</td>
<td>88</td>
<td>27</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>H35</td>
<td>25</td>
<td>68</td>
<td>20</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>H37</td>
<td>22</td>
<td>69</td>
<td>22</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>H38</td>
<td>24</td>
<td>84</td>
<td>26</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>H47</td>
<td>26</td>
<td>74</td>
<td>14</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>H48</td>
<td>28</td>
<td>82</td>
<td>20</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>H51</td>
<td>32</td>
<td>95</td>
<td>28</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>H52</td>
<td>27</td>
<td>90</td>
<td>28</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>H55</td>
<td>24</td>
<td>70</td>
<td>22</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>H57</td>
<td>26</td>
<td>80</td>
<td>18</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>H59</td>
<td>30</td>
<td>80</td>
<td>25</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>M</td>
<td>26</td>
<td>80</td>
<td>23</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>SD</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

| t       | -0.56     | -0.34     | 0.78      | -0.08       | 0.66      |
| P (1-tailed) | 0.29     | 0.37      | 0.22      | 0.47        | 0.26      |
of subjects would be required to obtain statistical significance.

Table 5 shows the number of practice trials, mean reinforcers earned per 1-hr session, and percent correct during fluency building for subjects in both experimental groups. Subjects in the x1.25 group received significantly more practice trials during fluency building than did subjects in the x2 group ($t_{[20]} = -6.33, p < .001$). This is an expected result, given that the celeration contingencies used were designed to produce different rates of acceleration. That is, because a x2 contingency reinforces faster celeration than does x1.25, response rate should reach the fluency aim more rapidly and, therefore, subjects should require fewer timings before reaching the fluency aim. This assumption was confirmed by the present data.

There was not a significant between-groups difference in reinforcers earned per hour during fluency building ($t_{[20]} = -1.37, p = .09$; see Table 5). The VI reinforcement schedule was designed to ensure that reinforcement density would not differ across subjects. The present results suggest that this goal was achieved. In addition, there were no between-groups differences in mean percent correct during fluency building ($t_{[20]} = -1.45, p = .08$; see Table 5).

Initially, a third experiment was planned which would test whether superior REAPS posttest scores were a function of the additional practice received by subjects in the x1.25 celeration group. This additional experiment would have controlled number of practice trials across groups while maintaining the x2 and x1.25 celeration contingencies. Given the present data, however, the additional experiment is not required. Even with significantly more practice trials, posttest performance of subjects in the x1.25 celeration group was no better than that of subjects in the x2 celeration group.

**Topography.** Topography of responding during fluency building was measured by assessing median latency, duration, and amplitude of vocal utterances. Inter-quartile ranges also
Table 5

Experiment 2: Practice Trials, Reinforcers per 1-hr session (Rein), and Percent Correct During Fluency Building.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Trials</th>
<th>Rein</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Celeration: x2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H28</td>
<td>931</td>
<td>60</td>
<td>0.95</td>
</tr>
<tr>
<td>H30</td>
<td>1478</td>
<td>50</td>
<td>0.93</td>
</tr>
<tr>
<td>H32</td>
<td>1542</td>
<td>46</td>
<td>0.93</td>
</tr>
<tr>
<td>H34</td>
<td>1507</td>
<td>61</td>
<td>0.93</td>
</tr>
<tr>
<td>H36</td>
<td>1182</td>
<td>50</td>
<td>0.93</td>
</tr>
<tr>
<td>H49</td>
<td>1518</td>
<td>66</td>
<td>0.98</td>
</tr>
<tr>
<td>H50</td>
<td>1275</td>
<td>58</td>
<td>0.86</td>
</tr>
<tr>
<td>H58</td>
<td>1140</td>
<td>68</td>
<td>0.94</td>
</tr>
<tr>
<td>H60</td>
<td>1348</td>
<td>64</td>
<td>0.93</td>
</tr>
<tr>
<td>H53</td>
<td>1712</td>
<td>39</td>
<td>0.94</td>
</tr>
<tr>
<td>H56</td>
<td>1041</td>
<td>60</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>1334</td>
<td>57</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>242</td>
<td>9</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Celeration: x1.25</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H31</td>
<td>2740</td>
<td>55</td>
<td>0.94</td>
</tr>
<tr>
<td>H33</td>
<td>2197</td>
<td>54</td>
<td>0.97</td>
</tr>
<tr>
<td>H35</td>
<td>3501</td>
<td>69</td>
<td>0.93</td>
</tr>
<tr>
<td>H37</td>
<td>3034</td>
<td>54</td>
<td>0.96</td>
</tr>
<tr>
<td>H38</td>
<td>4591</td>
<td>58</td>
<td>0.89</td>
</tr>
<tr>
<td>H47</td>
<td>3088</td>
<td>59</td>
<td>0.97</td>
</tr>
<tr>
<td>H48</td>
<td>2499</td>
<td>63</td>
<td>0.99</td>
</tr>
<tr>
<td>H51</td>
<td>1881</td>
<td>61</td>
<td>0.95</td>
</tr>
<tr>
<td>H52</td>
<td>2912</td>
<td>60</td>
<td>0.98</td>
</tr>
<tr>
<td>H55</td>
<td>2136</td>
<td>62</td>
<td>0.92</td>
</tr>
<tr>
<td>H57</td>
<td>1983</td>
<td>66</td>
<td>0.96</td>
</tr>
<tr>
<td>H59</td>
<td>3541</td>
<td>69</td>
<td>0.90</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>2842</td>
<td>61</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>786</td>
<td>5</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\[ t \] = -6.33, -1.37, -1.45

\[ p \text{ (1-tailed)} \] = 0.00, 0.09, 0.08
are reported to assess whether variability in any topographical measure changed across fluency-building timings, which might indicate some form of behavior streamlining. Medians are reported instead of means, so that outliers would not have disproportionate impact.

Median latency across fluency-building timings is shown for subjects in the x2 celeration group in Figures 19 and 20, and for subjects in the x1.25 group in Figures 21 and 22. Thick lines indicate median latencies. Thin lines above and below the thick line indicate the 25th (lower thin line) and 75th (upper thin line) percentiles. In general, the pattern of data for median latencies in these figures represents a negatively-accelerated decreasing function. That is, median latency decreases across timings throughout the experiment, but the amount of that decrement is less as the experiment progresses. This general pattern in response latencies also is seen in experiments on automaticity (Bloom, 1986).

Response stereotypy is another typical outcome found in automaticity experiments related to latency. Stereotypy (i.e., a decrease in variability of topographical measures) would be demonstrated in the present experiment if inter-quartile ranges decreased as the experiment progresses. Visual inspection of Figures 19 through 22 suggests that this may have happened with most subjects (H28, H32, H36, H49, and H56 in the x2 celeration group, and H31, H33, H35, H37, H38, H48, H55, H57, and H59 from the x1.25 celeration group).

It is not clear from these data whether decreased variability of latency is a function of (a) the general process of becoming fluent, or (b) the narrowing in the range of IRTs eligible for reinforcement. IRT ranges were determined based on a percentage of a target IRT. As the target IRT decreased, the absolute value of the range of IRTs eligible for reinforcement also decreased. Hence, because latency is a component of IRTs (the other component is utterance duration), it is possible that constriction of the IRT ranges produced less variable latencies, not fluency.
Figure 19. Experiment 2: Median latency (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
Figure 20. Experiment 2: Median latency (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
Figure 21. Experiment 2: Median latency (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
Figure 22. Experiment 2: Median latency (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
Although latency may be confounded with reinforcement, response duration and amplitude are not. Median utterance duration and amplitude are shown in Figure 23 for one representative subject (H35 and H50) from each celeration group. In general, data from Figure 23 are representative of the patterns obtained for all subjects in Experiment 2. That is, there is no trend in slope, and inter-quartile ranges do not show any obvious changes across timings. Hence, these topographical measures did not show any clear effects of fluency building. Duration and amplitude data for the remaining subjects in Experiment 2 are shown in Appendix C.

**Summary.** Efficiency data from the x2 celeration group were similar to data obtained during Experiment 1 under similar conditions. This finding demonstrates the reliability of the current procedures for obtaining efficient responding during fluency building for a x2 celeration contingency and 30 responses per minute fluency aim. Efficiency of responding during fluency building was not as strong, however, for subjects in the x1.25 celeration group. Responding of subjects in the x1.25 celeration group tracked the celeration line, but, toward the end of fluency building, did not stay within the programmed range of IRTs eligible for reinforcement. One potential explanation of this finding is that the IRT range for the x1.25 celeration contingency became so narrow toward the end of fluency building that subjects were unable to pace responding to stay within the range.

Even though subjects in the x1.25 celeration group had significantly more practice than did subjects in the x2 group, performance on REAPS posttests was not different. In addition, the fluency-building procedures employed in Experiment 2 affected response latency, but did not change utterance duration or amplitude. Median latencies decreased across fluency-building
timings. Inter-quartile ranges for the latency measure also became smaller across timings for most subjects, but the cause of this effect is uncertain.

Figure 23. Medians (thick lines) and inter-quartile ranges (thin lines) for utterance duration (left panels) and utterance amplitude (right panels) for two representative subjects from the x1.25 (top panels) and x2 (bottom panels) celeration groups.
GENERAL DISCUSSION

The present experiments had two objectives: (a) determine the appropriate fluency aim for reading binary numbers, and (b) assess whether two celeration rates commonly used in precision teaching have different effects on fluency. Both objectives were achieved. Results of Experiment 1 show that a fluency aim of 30 responses per minute at a x2 celeration rate produced the most effective and efficient responding. Results from Experiment 2 show that celeration rates of x2 and x1.25 did not differentially affect REAPS posttest performance. In addition, response duration and amplitude were unaffected by fluency building.

Interpretation of Results

Experiment 1: Selection of a fluency aim. Precision teachers can determine fluency aims based on responding of hundreds of students. In the present experiments, however, a fluency aim was determined with only 18 subjects. The current approach had the advantage that REAPS criteria were selected before the experiments, and each individual subject’s performance was empirically evaluated against those criteria. Although some practitioners have attempted to specify REAPS criteria rigorously (e.g., Johnson & Layng, 1992), the selection of fluency aims based on performance on REAPS posttests typically is not a systematic process (Robbins, 1994).

Even with the rigor used in the present experiments, however, not all REAPS criteria were useful. Perhaps some of the posttests were not sufficiently sensitive to detect differences between groups. It also is possible that posttests had low construct validity. That is, there may be a weak conceptual relation between the posttests and what they were designed to test. Each REAPS test is explored in turn.

Retention and endurance were the least useful REAPS tests because all subjects in Experiment 1 met the criteria for adequate performance on those tests. Furthermore, for all
subjects, scores on the retention test were higher than the mean of the last 5 timings during fluency building. Although two weeks is the standard retention interval in precision teaching, that may not be sufficiently long to provide a strong test. Perhaps one or two months would be better (Johnson & Layng, 1996).

Scores on 3-min endurance tests were similar to scores during 1-min fluency-building timings. Perhaps 5 min would provide a better measure of endurance. Stability and application were the most useful REAPS tests, because they showed that some subjects in the 25 responses per minute group in Experiment 1 were not fluent. Even though stability discriminated fluency, it might be a stronger test if distracting stimuli were auditory rather than visual. For example, random decimal numbers could be presented. One potential problem with that, however, is that auditory stimuli must not interfere with voice recognition. Presenting such stimuli through headphones would probably overcome that potential problem.

Similarly, even though application could discriminate between groups with different fluency aims, that test probably should be made more difficult to enhance sensitivity. For example, more challenging binary operations such as AND, OR, NOT, and XOR could be used. In the present application test, subjects manipulated binary numbers with only four or fewer digits. Perhaps larger binary numbers would make the test more difficult and therefore more sensitive.

Although the changes to REAPS tests listed above may improve the ability of that set of tests to detect differences between groups, it is uncertain whether those changes would improve the utility of all REAPS tests. It also is uncertain whether adduction provides a useful addition to that set of tests. Adduction was not systematically related to fluency aims or celeration rates, and it did not help to distinguish between the present groups. Furthermore, some subjects scored
high on the adduction posttest, whereas others were unable to answer any adduction questions correctly. Although the status of adduction remains uncertain, it is conceptually intriguing, and therefore should be included in future fluency studies after an appropriate functional definition is developed.

All of the REAPS criteria used in the present experiments have high face validity (i.e., they seem intuitively related to fluency). It is uncertain, however, which criteria are necessary and sufficient to produce fluency. Perhaps other predictors of fluency such as response topography will provide other or even better criteria. Regardless, it is time for precision teachers and researchers to broaden the search for fluency criteria, and test all candidates experimentally.

Precision teachers and researchers need a more systematic way of generating fluency criteria. Some people have suggested basing fluency aims on (a) performance of individuals who are expert performers on a particular skill, (b) a percentage of the speed that performers are able to obtain on a skill that is a component of the one being used for an experimental task, or (c) response rate on a task with a similar difficulty level (Binder et al., 1995). These methods may be easier than the method used in Experiment 1, but they forsake individualization, which is a hallmark of behavioral education. Compromise is often required in applied settings, but such a compromise may not be prudent. Regardless, the lack of standardized criteria for determining fluency is perhaps the most significant obstacle to a systematic fluency research paradigm.

**Experiment 2: Celeration.** The reinforcement schedule used in the present experiments effectively controlled celeration rates. This occurred despite the decline in response efficiency toward the end of Experiment 2 for subjects in the x1.25 group. Because celeration was controlled even though subjects earned less reinforcement, subjects were essentially responding according to the celeration target contingency despite being under extinction conditions.
Although subjects in the x1.25 group did not obtain all reinforcers in later timings, mean reinforcement rate for Experiment 2 was similar across celeration groups. For this to occur, the x1.25 celeration group must have been more efficient than the x2 group in the first part of fluency building. The IRT range for the x1.25 group did not change as much between timings as it did for the x2 group. It is therefore reasonable to assume that subjects in the x1.25 group were able to track the celeration contingency more easily at the beginning of the experiment than were subjects in the x2 group, because changes in the IRT range did not occur as rapidly.

The upper and lower limits of the IRT range were based on a percentage of the target IRT for a given timing. Target rates increased at a slower pace for the x1.25 group than for the x2 group. The absolute value of the IRT range was therefore small for the x1.25 group for a longer period toward the end of fluency building than it was for the x2 group. Perhaps tracking the reinforcement contingency became more difficult for subjects in the x1.25 group because of prolonged exposure to a reinforcement contingency with small IRT ranges.

The progressively narrowing IRT range also can explain the decrease in efficiency shown in Figures 6-9, 13, 14, 17, and 18. As the fluency-building condition progressed, the limits of the range of reinforceable IRTs not only decreased to produce an increase in rate, but the absolute value of the IRT range also decreased. It was therefore more difficult for IRTs to fall within the programmed range as the experiment progressed.

There are at least two changes that could be made to the present procedures to avoid the problems involved with using an IRT range that becomes smaller as fluency-building progresses. First, the IRT range could be based on a fixed interval above and below the target IRT instead of a percentage of the IRT. That way, the target IRT could change without changing the absolute value of the IRT range. This suggested procedure introduces an additional problem, however:
To maximize control of response rate in the first portion of fluency building, the IRT range would have to be sufficiently large so that subjects’ behavior would contact the experimental contingencies. As the target IRT decreases, however, the width of the IRT range would remain the same, so a wide range of IRTs would receive reinforcement later in fluency building. For example, if upper and lower limits of 2 s above and below the target IRT were used for the IRT range, the upper and lower limits of the IRT range for a target IRT of 6 s would be 8 s and 4 s, respectively. The same values for a target IRT of 2 s would be 4 s and 0 s. This is a floor effect. Using percentages of the target IRT avoids this problem.

To counteract a potential floor effect, a progressively decreasing absolute range could be used. That is, IRT ranges could remain fixed for a specified number of timings, then decrease as the target IRT decreases toward zero. For example, an IRT range of 2 s above and below the target IRT could be used for the first n fluency-building timings, then a range of 1.5 s above and below for the next n timings, and so on. The change in values above and below the target IRT, and the number of timings between modifications of the IRT range, could be manipulated experimentally. This procedure could determine the optimal arrangement to produce efficient responding that tracks the target rate and produces reinforcement for responding throughout the experiment. One potential disadvantage of this procedure is that different sets of parameters may need to be developed for different celeration rates, which may introduce an experimental confound.

Celeration values of x2 and x1.25 did not differentially affect fluency outcomes. That does not necessarily mean that celeration is not an important variable for producing fluent behavior. Celeration values greater or less than the values used in the present experiments should be tested to see if there are any performance benefits with celeration values greater than
x2, or problems with celeration values below x1.25. Additional experiments are required to
determine effects of other celeration values on fluency outcomes. To produce behavior that
tracks other celeration targets, it is necessary to monitor efficiency closely to ensure that subjects
are receiving reinforcers throughout fluency building. Regardless, results of the present
experiments are important for behavioral education, because x2 and x1.25 are commonly
considered by precision teachers to be optimal and minimal limits, respectively.

One potential criticism of the present experiments is that calendar time was not a part of
the way in which celeration was calculated. Some precision teachers maintain that celeration is a
frequency count of behavior per time, per time (Binder et al., 1995). That is, correct response
rate is measured across one-minute timings, then is graphed according to calendar time along the
horizontal axis on standard charts, then celeration is calculated. For precision teachers,
therefore, celeration is a measure of how response rate increases over calendar time.

Calendar time was not explicitly considered in the present experiments other than having
a standard number of practice timings each day and requiring subjects to participate Monday
through Friday. Hence, celeration was calculated based on the approximate number of timings
given per week by precision teachers (i.e., 20), even though that number of practice timings was
given daily instead of weekly. There is, however, a potential problem with the way that
precision teachers calculate celeration. Instead of plotting response rates on celeration charts for
each fluency-building timing, mean correct response rates for the day are plotted regardless of
the number of timings. Hence, the number of timings represented by each data point on a chart
used to calculate celeration might vary across days and students. This introduces a confound that
would be unacceptable in most experiments.
Precision teachers typically consider celeration to be a dependent variable. That is, celeration is monitored during fluency building to determine how learning is progressing, and interventions are introduced if celeration is low. Hence, for precision teachers, celeration simply is an indicator of learning. In the present experiments, however, celeration was treated as an independent variable. That is, celeration was manipulated experimentally to determine its effects on fluency. No effects were found. It is therefore possible that explicitly manipulating celeration diminishes its effectiveness. Perhaps celeration is not an independent variable, and only affects fluency when it is allowed to progress naturally. If that is correct, then the practice of precision teaching, particularly within generative instruction, will need to be changed significantly.

Topography

Other than latency, topography of vocal responses was unaffected by fluency-building procedures. The simplest explanation for why latency decreased for all subjects in Experiment 2 is that latencies needed to decrease for responding to be reinforced. Reinforcement was based on response IRT, however, not latency. Latency is only one component of an IRT; the other component is utterance duration. Hence, it might be expected that both latency and duration should decrease if IRTs decrease. This did not occur in the present experiments, however. Response duration remained relatively unchanged throughout fluency building.

Variability of response latencies decreased across fluency-building timings for most subjects in Experiment 2. It is uncertain why this occurred. Decreased variability in response latencies also has been found in a study of automaticity (Carlson, Sullivan, & Schneider, 1989). Subjects in that study were instructed to practice solving logic problems with simple electronic circuits for a certain number of practice trials. Neither response rates nor response latencies were
explicitly reinforced in that study. Nonetheless, because of the decrease in variability of response latencies, subjects’ behavior was described as stereotyped and automatic. Perhaps a similar process was involved in the present experiments.

**Vocal operants.** Operant research on vocal behavior may shed some light on the present results in the area of response topography. Several vocal-operant response classes were studied during the 1950’s and 1960’s (Azrin, Jones, & Flye, 1968; Flanagan, Goldiamond, & Azrin, 1958; Salzinger, Waller, & Jackson, 1962; Starkweather, 1960; Wilson & Verplanck, 1956). Experimenters have demonstrated operant control of vocal behavior with chickens, parakeets, mynah birds, guinea pigs, sea lions, and dolphins (Hake & Mabry, 1979), cats (Molliver, 1963), monkeys (Leander, Milan, Jasper, & Keaton, 1972), and human infants (Routh, 1969), and have brought under experimental control pitch (Lane & Shinkman, 1963; Manabe, Kawashima, & Staddon, 1995), latency (Cross & Lane, 1962), amplitude (Lane & Shinkman, 1963), duration (Lane, 1964), and magnitude (Lane, 1960) of vocal responses.

Cross and Lane (1962) reported that latency could be manipulated independently from amplitude. Lane (1960) reported similar findings with response duration. It is possible that the present experimental procedures affected latency independently of the other topographical measures because of the way in which reinforcement was arranged. Further research needs to be conducted to determine whether present changes in response topography are due to effects of fluency, or some other operation related to the IRT contingency.

**Behavior streamlining.** Present results neither support nor disconfirm the notion of behavior streamlining. Response latencies decreased during fluency building, but it is uncertain whether that was because of streamlining. It is possible that the critical aspects of topography related to streamlining of reading binary numbers (e.g., covert responses such as counting place
values) were not measured using present procedures. In fact, behavior streamlining may not be a useful concept unless topographies are measurable (i.e., not cognitive behavior). Perhaps future studies of streamlining should be based on a motor task instead of tasks that involve cognitive behavior.

Experimental Procedures

 Benefits. The celeration contingency used in the present experiments allowed response rate to be controlled precisely while maintaining a similar reinforcement rate between experimental groups. This permitted an experimental assessment of celeration that was untainted by the extraneous variables typically found in a precision-teaching classroom, such as the social contingencies involved when peers observe a student’s practice timings during fluency building.

 Although the main analyses in the present experiments involved between-groups comparisons, single-subject data also were collected. In particular, efficiency data were useful in selecting a fluency aim in Experiment 1. In addition, data from topographical measures for individual subjects were useful for determining whether fluency affects response topography as fluency building progressed. Such single-subject analyses provide experimenters with valuable information about how individuals respond to fluency-building procedures.

 Potential changes. Despite the advantages of the present experimental procedures, they could be refined to enhance the experimental analysis of fluent behavior. Specifically, improvements could be made to a few of the fluency-building procedures used in the present studies. For example, it is uncertain what effect computerized feedback graphs that were presented between fluency building timings may have had on the effectiveness of the reinforcement contingency. In future studies, feedback graphs should be eliminated to assess whether celeration could be controlled without the graphs. If it is determined that feedback
graphs are an essential element of fluency procedures, perhaps the instruction given to subjects to keep response rate near the celeration line on the graph could be eliminated to see if celeration can be controlled without it. In addition, the celeration line itself could be eliminated. It is possible that feedback graphs controlled celeration, not the reinforcement procedure. That possibility also should be investigated to determine the optimal conditions for controlling celeration.

Several changes also could be made to the reinforcement procedures designed to control celeration. For example, the initial target rate was fixed for all subjects at 10 responses per minute. Perhaps target rate could be determined for each subject based on rate during the first fluency-building timing, or the average response rate of the first 5 timings. This has the advantage of accommodating individual differences between subjects. It also has the disadvantage that subjects are likely to receive a different number of practice timings within a group because of different initial target rates, which may be experimentally undesirable.

The celeration contingency was designed specifically to control response rate. This leads to the criticism that rate was artificially constrained, and that subjects’ posttest performance might have shown differences if allowed to accelerate naturally. Future studies should include an additional group of subjects who were exposed to the VI contingency, with and without upper and lower limits on the IRT range, to test this proposition. As was mentioned previously, another improvement to the experimental procedures would be changing the way in which the IRT range is determined, so that efficiency of responding also is maximized.

**Voice Recognition Apparatus**

Advances in voice recognition technology are occurring rapidly. Even so, very few studies of human operant behavior have taken advantage of voice recognition (cf. Baron &
Journey, 1989; Beins, 1989; DeHaemer, Wright, & Dillon, 1994). Early studies of vocal operants with humans used voice-operated relay racks (e.g., Cross & Lane, 1962) which could measure various aspects of utterance topography, yet were unable to discriminate different words reliably. To be useful for studying vocal-operant behavior, an apparatus should be able to reliably detect the vocal response being studied, and should provide a link between the subjects’ speech and an automatic recording and programming apparatus (Shearn, Sprague, & Rosenzweig, 1961). The voice-recognition apparatus used in the present experiments satisfied these requirements.

Accuracy of voice recognition requires a location for the experiment that is very quiet, because the voice-recognition apparatus requires a sound that is 25 dB above ambient room noise before an utterance is detected. If ambient noise levels are high, the sound pressure levels required to produce an additional 25 dB of noise are much higher than if ambient noise is low. The additional effort required to produce higher sound pressure levels might add an undesired response-cost component to an experimental situation (cf. Davison & Kirkwood, 1968).

Although voice recognition was useful in the present experiments, it is not essential for studying fluency. Subjects could have made responses using the computer keyboard, which also would permit measurement of response latency. One disadvantage of using the keyboard, however, is that many subjects are not proficient at typing, and that would inhibit fluency.

Future Studies

The present methods were effective, and provide a good starting point for experimental fluency research. There are at least two obvious areas of research suggested by the present results. First, the tests for REAPS outcomes need to be refined such that they readily provide information about when subjects become fluent. In addition, different celeration values should
be tested to see if there are benefits of celerations greater than x2 or problems obtaining fluent performance with celerations below x1.25.

Adduction posttests can be included in future studies if (a) their conceptual relation to fluency can be clarified and (b) a useful functional definition of adduction is developed. One possible means of accomplishing this task would be to identify two or more experimental tasks that are components of a composite skill, then give subjects practice with both tasks until they are fluent, and then test for adduction. For example, precision teachers have used mathematical operations that require multiple repertoires such as algebra to study fluency (Miller & Heward, 1992).

Although present experiments provided no support for behavioral streamlining, they were not designed to study it. To study streamlining, an experimental task is required that has been shown to have observable changes in topography as responding becomes fluent (e.g., motor responses). For example, a task such as assembling complex mechanical equipment could be used, and topographical aspects of hand position could be measured to assess streamlining. Motion-sensing devices already have been developed for use in virtual reality environments, and could be used in such an experiment.

Celeration is only one of many variables that may be critical to the development of fluency. Precision teachers have identified what they believe to be the critical set of fluency-building procedures, but they are relying on future laboratory research to confirm their assumptions (Binder et al., 1995). For example, Lindsley (1996a) contends that high error rates at the beginning of fluency building are desirable. It is uncertain whether skills need to be trained to high accuracy levels prior to starting fluency building. This proposition has not been tested experimentally. Other variables that may be critical to fluency include (a) analysis of
component and composite skills, (b) sequencing of materials in a generative-instruction curriculum, (c) graphical feedback to learners about fluency-building performance, and (d) controlled- versus self-pacing. Further research can be conducted by manipulating these variables individually and in various combinations in an effort to determine the critical set of variables necessary to build fluent behavior.

**Applied and Social Validity**

The implications of the present research for precision teaching are significant. The results of Experiment 1 underscore the need to develop a more precise functional definition of fluency, such that it is possible to specify a general set of procedures for generating fluency in the classroom. The definition should allow precision teachers to generate a variety of fluent behavior with different difficulty levels and modalities (e.g., vocal versus manual responses), and should account for behavior at different levels within a hierarchy of learning objectives. Furthermore, procedures to determine fluency criteria based on REAPS outcomes need to be specified precisely and tested extensively with a variety of academic tasks with varying difficulty levels.

The results of Experiment 2 have important implications for how precision teachers arrange fluency-building procedures. If learners can perform equally well on tests of fluency outcomes with celeration rates of $x^2$ and $x^{1.25}$, precision teachers should attempt to allow learners’ behavior to accelerate naturally within these parameters. It is an empirical question whether arranging conditions such that a particular celeration rate is achieved is the best method, or whether allowing celeration to occur naturally is more appropriate. Nonetheless, based on my observation of pacing procedures used at the Morningside Academy, it seems likely that learners will report higher satisfaction when they are allowed to proceed at their own pace.
Individualization of learning always has been a hallmark of behavioral education (Skinner, 1954). The present results provide additional empirical support for this tradition in both the classroom and the laboratory.

Conclusions

Even though precision teaching and generative instruction have promoted the benefits of fluency for years, the experimental analysis of fluency is in its infancy. Because the notion of fluency originated in the classroom and not in the laboratory, researchers need to identify the variables of which fluency is a function, while eliminating extraneous variables. The most important initial step in the development of a fluency paradigm is to determine a functional definition of fluency based on replicable procedures. The present experiments provide some procedural innovations that may eventually lead to a comprehensive approach to the study of behavioral fluency.

To strengthen the fluency paradigm, a standard set of procedures should be developed that is similar to those used in the experimental analysis of operant behavior with non-humans (cf., Sidman 1960). The elegant simplicity of the rat’s bar press or the pigeon’s key peck has allowed operant researchers to build a vast system of knowledge about a variety of basic behavioral principles; the study of fluency needs a similar foundation. Although the study of fluency should undoubtedly be continued in the laboratory, applied researchers and practitioners can contribute by identifying variables related to fluency that can be studied in the laboratory, and by implementing experimental findings in systematic ways to extend the generality of laboratory discoveries.

Behavior analysts occasionally use colloquial terms to describe fundamental behavioral concepts (e.g., “reinforcement”) when an obvious relation exists between the term’s colloquial
meaning and a behavioral process they are trying to describe. This certainly is the case with the term “fluency.” Reinforcement is a behavioral primitive (i.e., a basic fundamental behavioral process) that is easy to identify procedurally, whereas fluency is not easily identified. Functional REAPS criteria may be useful as indicators of fluency, but no comprehensive definition of fluency itself exists. Furthermore, it is uncertain which of the REAPS criteria are necessary for fluency.

Behavior analysis is an inductive science, in that systems of knowledge are developed by trying to understand the most basic concepts first, then relating those concepts to form more complex principles. Unfortunately, the way that behavioral educators describe fluency violates that scientific approach. Fluency is described as an intervening variable without obvious conceptual or methodological roots. Nonetheless, procedures such as fluency may help to achieve the great promise of behavioral education. What greater legacy could Skinner have than providing the foundation for a technology that makes education effective, efficient, and satisfying to its consumers?
REFERENCES


White, O. R. (1986). Precision teaching - precision learning. Special Issue: In search of excellence: Instruction that works in special education classrooms. Exceptional Children, 52(6), 522-534.


APPENDIX A

Subject Recruitment Poster

Subject Recruitment Sign-up

Fluency with Number Systems on a Computer: Study #97-SU-54

Participants are needed for an interesting study to be conducted in the Department of Psychology. Only persons who are not Computer Science majors or minors and who obtain a particular range of scores on a pretest will be accepted for participation. The experiment will be conducted in a temperature-controlled laboratory in the Oglebay Annex during the spring 1998 semester. Participants have the opportunity to earn approximately $5 per hour of participation, which also includes a $20 attendance bonus for coming to all scheduled sessions (up to 10 1-hr sessions).

Please call 293-2001 extension 857 and leave a message on the voice mail system for Kevin Munson with your name, phone number, and the best time to call. This will indicate your interest in participating in this experiment. You will be contacted within one week.
APPENDIX B

Programmed Instruction Sequence

Set 1: Converting Decimal and Binary numbers, 37 Questions
Q. The number system most commonly used by people (the decimal number system) is based on multiples of the number ______.
A. ten (10)

Q. In other words, the decimal number system has a b______ of 10.
A. (b)ase

Q. The base of a number system refers to the number of digits used in that system. Hence, the decimal number system has a total of ______ digits.
A. 10 (ten) [ 0,1,2,3,4,5,6,7,8,9 ]

Q. In the decimal number system, the value of a number (for example, 569) is computed by considering each individual ______ in the number (that is, 5, 6, and 9).
A. digit

Q. We also consider the column position of each digit relative to the column on the far right. A digit one column from the right is in Column 1. A digit two columns from the right is in Column 2. A digit in the right column is in Column ______.
A. 0 (zero)

Q. If the digit 7 is in Column 0, it represents ______ decimal "units."
A. seven (7)

Q. If the digit 2 is in Column 1, it represents ______ decimal units.
A. twenty [ 2 x 10 ]

Q. If the digit 4 is in Column 3, it represents ______ _____ decimal units.
A. four thousand [ 4 x 10 x 10 x 10 ]

Q. For each column we examine from the farthest right column (i.e., each time we move left a column), the value of the digit is multiplied by one more ______.
A. ten (10)

Q. In the number 500, the 5 is in Column 2, so it represents five ______ units (that is, 5 x _____, x ______).
A. hundred 10 10

Q. The number 649 represents six ______ (Column 2) + ______ (Column 1) + ______ (Column 0).
A. hundred [6 x 10 x 10], forty [4 x 10], nine
Q. In any number system, to calculate the value of a number, we need to know three things: (a) the _______s in the number, (b) the _______ positions of these _______, and (c) the _______ of the number system.
A. digit(s) column digits base

Q. The _______ number system is a convenient number system for humans because we have 10 fingers.
A. decimal

Q. Computers do not have fingers. Instead, they calculate by monitoring the presence and absence of electric current in circuits. Hence, the number system used by computers has only two d_______.
A. (d)igits

Q. The number system used by computers has a base of _______.
A. two (2)

Q. Computers can perform millions of calculations every second because e______ c______ can be turned on and off very quickly.
A. (e)lectric (c)urrent

Q. Hence, computers achieve their phenomenal speed because all calculations are based on the presence or absence of _______ _______, and a number system with a base of _______.
A. electric current two

Q. Because the number system used by computers has only two states, it is called _______nary. (from the Latin for two components)
A. bi(nary)

Q. In the binary number system, the digits 1 and 0 refer to the states of electric current, _______ and _______, respectively.
A. on off  [ in that order ]

Q. In the decimal system there are 10 digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9). In the binary system, there are _______ digits, namely _______ and _______.
A. two (2) 0 1

Q. The binary number system has ______ as its base.
A. two (2)

Q. To calculate the value of a number in the binary system, we consider the _______s (0,1), the _______ location of these _______, and the _______ of the binary number system (i.e., two).
A. digits column digits base
Q. If the binary digit 1 is in Column 0, it represents ______ binary "unit(s)."
A. one (1)

Q. If the binary digit 1 is in column 2, it represents ______ binary unit(s).
A. four [ $1 \times 2 \times 2$ ]

Q. The binary number 1000 represents ______ units.
A. eight [ $1 \times 2 \times 2 \times 2$ ]

Q. The technique for calculating the value of a binary number is the same as for calculating the value of a decimal number: Consider the digits, consider their column locations, and consider the base. The binary number ______ represents five units. ([ $1 \times 2 \times 2$ ] + [ $0 \times 2$ ] + 1)
A. 101

Q. In place of the phrase "the decimal number 9", we can use the expression 9D. The D after a number shows that the number is from the ______ number system.
A. decimal

Q. Similarly, 111B means that the number 111 is from the ______ number system.
A. binary

Q. Because we are considering two number systems in this lesson, it is important to specify which number system is being used. For example, 11 has the value of ______ if it is binary, and ______ if it is decimal.
A. three eleven

Q. The number 10B has only one value: ______.
A. two

Q. The number 101D has only one value: ______.
A. one hundred and one

Q. The number 111B has only one value: ______.
A. seven

Q. In the remaining questions, numbers expressed in words are decimal, and numbers expressed as numerals are from the number system represented by the letter to the right of the number (i.e., B, or D). For example, ______ (the number expressed in words) has the same value as 11D, but not 11B.
A. eleven

Q. Which has the larger value: 10000B, or 14D?
A. 10000 B [ $1 \times 2 \times 2 \times 2 \times 2 = sixteen$ ]
Q. Which has the larger value: 1000B, or 10D?
A. 10 D [ 1000 Binary = 1 x 2 x 2 x 2 = eight ]

Q. Which has the larger value: 1011B, or 10D?
A. 1011 B [ 8 + 0 + 2 + 1 = 11 ]

Q. To calculate the value of a binary number, consider three things: the digits, the ______ locations of these digits, and the base of the number system (______).
A. column two
1. Median utterance duration (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
2. Median utterance duration (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
3. Median utterance duration (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
4. Median utterance duration (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
5. Median utterance amplitude (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
6. Median utterance amplitude (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x1.25 celeration group.
7. Median utterance amplitude (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
8. Median utterance amplitude (thick lines) and inter-quartile ranges (thin lines) across 1-min timings for subjects in the x2 celeration group.
ABSTRACT

EFFECTS OF Celeration RATE ON BEHAVIORAL FLUENCY

By Kevin J. Munson, M.A.

In precision teaching, it has been argued that learning should continue until responding is fluent (i.e., it can be performed accurately and rapidly in the presence of distracting stimuli, even after a delay since its acquisition). Furthermore, it has been argued that the speed with which learners achieve fluency (i.e., celeration) also is important. The purpose of the present experiments was to test those propositions. Subjects read binary numbers on a computer screen, and spoke the decimal equivalent of those numbers into voice-recognition equipment. In Experiment 1, responding was reinforced with money arranged on VI schedules for correct responses with inter-response times within a programmed range. This procedure allowed response rate and celeration rate to be controlled precisely while maintaining a similar reinforcement rate across conditions. Each group of subjects received one of the following terminal-rate conditions: 25, 30, 35, and 40 responses per minute. The terminal rate of 30 responses per minute was the most effective and efficient. In Experiment 2, each group of subjects received a condition in which the target rate increased by either 100% or 25% every 20 sessions until responding reached 30 responses per minute. In precision teaching, 100% increases after 20 sessions are considered optimal, and 25% increases are considered minimal. No between-groups differences were found on fluency tests. Latency of vocal responses decreased, but other topographical measures did not change. Present results suggest that terminal response is an important correlate of learning, but celeration rate is not.