RETENTION, ENDURANCE, STABILITY, AND APPLICATION OF LEARNED PERFORMANCES AS A FUNCTION OF TRAINING CONDITION

Jason Cohen, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2008

APPROVED:

Sigrid Glenn, Major Professor
Janet Ellis, Committee Member
Michael Fabrizio, Committee Member
Richard Smith, Chair of the Department of Behavior Analysis
Thomas Evenson, Dean of the College of Public Affairs and Community Service
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies

Cohen, Jason. Retention, Endurance, Stability, and Application of Learned Performances as a Function of Training Condition. Master of Science (Behavior Analysis), December 2008, 39 pp., 4 tables, 16 figures, references, 17 titles.

A functional definition of fluency describes performance frequency ranges that predict retention, endurance, stability, application, and adduction as outcomes of practice. This experiment assessed these outcomes after different training conditions using a within-subject design. Participants in an experimental group learned new skills in a condition with rate and accuracy criteria, then in a yoked, rate-controlled condition with the same number of prompted responses and correct trials in practice. Control group participants received training in consecutive conditions with rate and accuracy criteria. Performance of individuals in the control group demonstrated practice effects. Data obtained from participants in the experimental group showed similar performance across conditions. Considering efficiency, the condition with rate and accuracy criteria was superior.

Copyright 2008

by

Jason Cohen

ACKNOWLEDGEMENTS

I would like to thank Dr. Sigrid Glenn for the guidance and feedback that she so masterfully provides. I would also like to thank the one-of-a-kind Dr. Janet Ellis who dedicates much of her time and energy to her students. A special thanks to Bryan Lovelace whose programming skills were instrumental in completing this project. My wife, Kathleen, deserves my appreciation for her patience and advice.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

Fluency is functionally defined as performance at frequency ranges that predict significant learning outcomes (Binder, 1996: Johnson & Layng, 1996). Qualitative descriptions include fast and accurate, effortless, flowing, automatic. Although technical language is not always consistent and empirical validation is weak, fluency is the definition of “true mastery” (Binder, 1988).

Several terms and acronyms associated with fluency have been introduced in the last few decades and terminology has evolved as research and practice inform the growing literature on the topic. Retention, maintenance, endurance, stability, application, adduction, and performance standards have been elements in well established and recent acronyms such as REAPS (Haughton, 1980), RESAA (Johnson & Layng, 1992, 1994, 1996), and MESAA (Johnson, 2007) For the purposes of this paper, the acronym RESAA as defined by Johnson and Layng (1996) will be used. For a history of the development of these terms, see Johnson and Layng (1996) and Binder (1996).

Johnson and Layng (1996) define the outcomes of fluency as follows. Retention is shown when a skill is performed at or near a previously attained rate after several weeks without practice. Endurance refers to performance rates sustaining through durations of time that extend past that of typical training. Performance shows stability when rates are near training criteria in the presence of distractions. Application refers to when someone applies previously trained component skills when learning a new composite skill. Adduction, a type of application, refers to when contingencies recruit the combination of trained component skills with almost no instruction. Johnson (2007) further clarifies that “adduction can occur only once, but the adduced repertoire becomes part of a behavioral history that can then be ‘applied’ over and over” (p.14).

As evidenced by Doughty, Chase, and O’Shields (2004) and subsequent replies by Binder (2004) and Kubina (2005), the acronyms and terminology are not used consistently within behavior analytic literature. However, advances in the literature continue with specification of terms in research, reviews, and technical writings. For example, Johnson (2007) introduced maintenance in place of retention and the previously mentioned distinction between application and adduction.

Terminology changes as a result of developing assessment and training procedures. Specific, standardized, training and assessment procedures do not exist within fluency-based instruction, but guidelines direct current practices. These guidelines have been continually refined through thousands of hours of application since the 1970’s. Several textbooks, dozens of articles in peer-reviewed journals, and an unspecified number of experiments by teachers and researchers also contribute to current practices.

As stated in the definition, performance is labeled fluent after meeting certain criteria on several assessments (RESAA). Neither standard assessments nor criteria exist, although examples and guidelines can be found in the literature as well as in practice. Retention is usually assessed after 4 or more weeks without practice. Durations 3 or more times longer than that of practice timings can be used in an endurance test. Stability tests include distractions, and Binder (2004) notes that “original use of the term *distractibility* referred to the influence of potentially competing stimulus control" (p. 284). Application tests may assess performance on a complex skill, of which previously trained skills are components. Acquisition of a composite is faster when performance of component skills is fluent. If a composite skill occurs with little or no instruction as a function of previously trained component skills, adduction is said to have occurred.

As operant conditioning procedures and measurement of the free operants of animals developed in the lab, Ogden Lindsley, one of Skinner's students, is recognized as the first to use these procedures to measure human behavior in applied settings (Binder, 1996). Precision teaching was later developed by Lindsley and colleagues, and techniques became more sophisticated and widely used. Identification of significant outcomes resulting from training skills to specific frequency ranges followed. Soon teachers and students began building response rates of academic skills as the efficacy of measuring frequency as a criterion for mastery became evident (Binder, 1996).

Whether referring to responses that are chained or emitted individually, speed is a feature of proficient performance of any skill. It seems logical, then, to target increases of response rate in practice or training procedures. It would be hard to argue that slower performance is more desirable than faster performance of equal accuracy, unless low-rate performance itself is targeted. Pauses between responses and long latencies between stimulus presentation and response are rarely, if ever, desired outcomes of practice. However, few educators take speed into account when teaching new skills.

When acquiring a new skill, a learner may receive several prompts and error corrections and practice multiple times before the skill is considered mastered. Outside of a definition of fluency, mastery can be described as a probability that a specific response will be emitted in a given situation. However, there is no quantifiable dimension of behavior related to probability (Johnston & Pennypacker, 1993). Rate, or frequency, of responses is the measure that fluency-based instruction uses to determine whether a skill has been mastered.

"As the principal datum of instructional behavior, frequency is fundamental to its methods" (Barrett, 2002, p. 74). And since a measure of frequency is not complete without a time referent, frequency per unit of time is a necessary dimension of behavior if any qualifying statements are to be made (Johnston & Pennypacker, 1993). This provides information on which performance decisions can be made; decisions not possible with a ratio or percent correct, a "dimensionless quantity" (Johnston & Pennypacker, 1993, p.103).

Although rate-building is the hallmark of fluency-based instruction, Kubina (2005) noted that "almost any practiced activity can qualify as rate building…because practice minimally consists of repeating behaviors in time" (p. 75). As fluent performance results from sufficient practice, rate-building procedures are used "as an efficient method for practice" (p.75).

The efficacy of rate-building procedures has been demonstrated in academic, professional, and clinical settings since their first applications. However, few controlled experiments have assessed the effects of rate-building procedures under RESAA conditions within the context of experimental questions assessing possible functional relations between behavioral rate and fluency (Doughty et al., 2004). Although the (RESAA) effects have been shown after rate-building procedures, the question remains as to whether or not the rate (count/minute) itself accounts for the effects that define fluency or if the rate is a result of something else involved in the procedures.

Since the effectiveness of rate-building procedures is established, further experimental assessment may not be as valuable to some educators as effecting behavior change. Also, rate-building procedures and effects have been identified inductively from practices in classrooms rather than tightly controlled, experimental settings.

> Consequently, the rate of displaying experimental control required of behavioral applications has become correspondingly less than the standards typical of laboratory research. This is not because the applier is an easy-going, liberal, or generous fellow, but because society rarely will allow its important behaviors, in their correspondingly important settings, to be manipulated repeatedly for the merely logical comfort of a scientifically sceptical audience. (Baer et al., 1968, p.92)

There are also methodological difficulties regarding experiments that compare procedures with and without rate criteria. Procedures that target skill acquisition imply permanent changes in subjects' repertoires, which makes controlling for sequence effects an issue in experiments that utilize within-subject designs. Other difficulties include controlling for practice effects, reinforcement, and inconsistency of relevant terminology and procedures relating to acquisition and error correction.

Practice effects can be controlled by yoking the number of trials or the amount of time spent practicing across conditions with and without rate criteria. Binder (2004) noted that "Controlling for amount of reinforcement… number of reinforcements and sources of reinforcement can be difficult to define unambiguously" (p.285). Reinforcement refers to a specific functional relation rather than just the presence of a preferred or pleasurable event such as praise. Procedures that include delivery of stimuli such as praise and attention should be described as reinforcement only if they function to increase the frequency of the behavior on which they are contingent. The reinforcers that increase the probability of accurate responses in rate-building procedures are the stimuli of interest in experiments that attempt to control for reinforcement effects. Also, it should be noted that precise language is a necessary condition for systematic replications and inductively identified principles (Johnston & Pennypacker, 1993). Inconsistent use of the terms stability, application, and adduction further complicate fluency research.

The scarcity of controlled research on the possible functional relation between fluency and response rate is evidenced in a recent review by Doughty et al. (2004). The authors evaluated whether empirical articles controlled for number of practice trials and reinforcement rate and if RESAA outcomes were assessed. Of the 48, peer-reviewed publications that fit the authors' search criteria, 29 were determined empirical. Of these empirical publications, 10 assessed retention and 4 tested endurance. Ten studies were determined to have assessed extension, a category that included stability, application, and adduction; however, interobserver agreement was 71%.

Controlled experiments

Doughty et al. (2004) identified three articles that controlled for practice trials and reinforcement and assessed fluency-related outcomes (Evans & Evans, 1985; Evans Merger, & Evans, 1983; Shirley & Pennypacker, 1994). Evans et al. (1983) assigned participants to high, medium, and low rate groups depending on baseline measurements and yoked performance of all participants to 940 trials of saying letter sounds. Participants practiced during three, 1-min timings each day until they met the rate criterion of 5 or fewer errors. Praise and feedback were delivered to participants after each timing. If they met the mastery criteria in fewer than 940 trials, response rate was held constant with a controlled reader. Post-tests conducted upon completion of training procedures consisted of an untimed presentation of 100 CVC trigrams. Results on the post-tests indicated that participants in the high-rate group attained greater percent correct scores than those in other groups. Doughty et al. (2004) did not identify this study as one that tested extension (stability, application, or adduction). However, it could be argued that the post-test assessed performance in an application or adduction condition since component skills were trained and composite skill performance was tested.

Evans and Evans (1985) conducted a follow-up study with procedures similar to those of Evans et al. (1983), but rate criteria for groups were 60, 90, and 120 responses/min. Results on post-tests were median rate changes and favored the medium rate group.

However, those studies, do not answer experimental questions regarding the possible relation between response rate and fluency if a component of such experimental procedures involves “separating the effects of practice and rate building” (Doughty et at., 2004, p. 18). In these studies, performance was initially practiced without constraints, although it was eventually constrained by the controlled readers. The authors controlled rate of reinforcement by delivering praise and feedback after timings, but they did not account for the number of error corrections, prompts, or other interactions between timings that may have included further praise or feedback.

Shirley and Pennypacker (1994) yoked the number of times participants practiced spelling 10 words on each of two different word lists. Practice on each list continued until accuracy on both lists was 100% and the rate of correct letters written reached a previously determined criterion. (This criterion for each participant, which was assessed in a baseline condition, equaled the rate of correctly written letters of their names.) This phase lasted 4 weeks, and spelling performance on retention tests was assessed 10 days after completing mastery criteria for each list. Rates and ratios of correctly written letters were slightly higher in post-tests on word lists in which there were rate and accuracy criteria. Performance rates were assessed during and after training conditions without a rate criterion, which is consistent with one of Binder's (2004) recommendations for research that compares rate-building and rate-controlled procedures.

Again, no participants practiced in a condition in which accuracy was the only criterion. Therefore, it is not possible to conclude that a rate criterion results in fluent performance if yoked trials were not practiced in a rate constrained context.

According to Doughty et al. (2004) the three previously mentioned publications "represent the best methods used to date to control for practice and rate of reinforcement" (p. 16). If the control of these two variables as described by the authors is critical in fluency research, then future experiments may yoke the number of practice trials and reinforced responses to better control for practice and reinforcement effects. Wheetley (2005), in an unpublished master's thesis, did just that.

Participants received instruction and practice on matching English letters and Kanji symbols on a computer touch screen after a pre-test determined any biases of letter to symbol associations. Within subject, ABA and BAB designs were used to assess the effects of training procedures with rate and accuracy criteria (R+Ac) and without a rate criterion (Ac). The number of error correction (including initial acquisition) trials and the number of correct practice trials were the same in both conditions. Performance was assessed in endurance and stability conditions after participants completed seven or more correct responses and zero errors in a 20s (R+Ac) drill, or seven correct responses and zero errors in a seven-trial (Ac) drill (which were 100% accuracy criteria of those conditions). Retention, endurance, stability, and application conditions were conducted several days after fulfilling the mastery requirements of each training session.

Results indicated that performance trained in R+Ac conditions resulted in slightly higher accurate rates during tests relative to the yoked Ac conditions. Of the 20 tests on which within-subject comparisons of yoked conditions can be made, 15 favored the R+Ac conditions. Stability tests possibly confounded these general findings in that the duration of tests was 3 times that of practice drills, and the distracter involved the experimenter walking in and out of the room, which may have varied across tests. Another issue regarding the overall results is that 1 participant in the yoked Ac condition did not acquire the skill. This participant continued to make errors without corrections due to the yoking procedure.

The current experiment is an extension of Wheetley (2005) with several procedural differences. A pre-training phase was conducted in order to familiarize participants with experimental procedures without the opportunity to practice errors. The number of stimuli in each training condition was reduced to six. This allowed for more performance comparisons as two Ac conditions could be yoked to two R+Ac conditions. Control subjects were added to address sequence effects, and testing procedures were modified. These differences aside, the experimental question remained the same. The current experiment assessed endurance, stability, and application outcomes after R+Ac training and Ac training conditions in which the number of correct practice trials and reinforced responses were yoked.

CHAPTER 2

METHOD

Participants and Setting

Five undergraduate students with no prior knowledge of the Japanese language participated in this experiment. Participants were paid $7/session with a $25 bonus upon completion. All sessions were conducted in a small room containing a chair, a computer, a table with a monitor and a touchscreen.

Response Definition, Apparatus, Stimuli and Data Collection

Finger presses on one of six comparison stimuli in the presence of a sample stimulus on a computer touch screen were recorded. All trials in training and test phases were completed on a Keytech Magictouch USB-X® touch-screen that covered the entire 17-in. monitor of the desktop computer. The experimenter calibrated the touch-screen prior to each session. There were four sets of 12 stimuli in this experiment. Each of the sets contained 6 English letters and 6 Kanji symbols. In addition, a set of pre-training stimuli contained 6 geometric shapes and 6 cardinal numbers. At the beginning of each participant's first session a randomly assigned sequence of the four sets was determined by the experimenter. Stimuli were arranged with the sample stimulus in the center of the screen and the six comparisons distributed around the sample. At least 2 in. separated each stimulus from all others (Figure 1). Programs developed in Microsoft Visual Studio® collected all data.

Experimental Design

A single-subject A-B-A-B design was used to compare the effects of two types of training, designated as Rate plus Accuracy (R+Ac) and Accuracy Only (Ac), on four tests. A pre-training session was used to familiarize the participants with the task prior to training. Following pre-training 3 participants received training in conditions R+Ac, Ac, R+Ac, Ac. Two additional participants received repeated training in the R+Ac condition to assess practice effects of repeated R+Ac training.

R+Ac conditions included training criteria for both accuracy and rate, whereas the criteria for completion of the Ac condition consisted of completing training components yoked to the number of practice trials and number of instruction trials in the previous R+Ac condition. The effects of these training methods were tested in conditions that have been said to show "fluency" (endurance, stability, application). Rate tests were also conducted.

Procedure

Training

At the beginning of each session participants received general information describing instruction and practice procedures, including the mastery criteria for that condition. Training consisted of two components: instruction and practice. Each condition began with a 6-trial instruction exercise that prompted correct responses to a new set of stimuli by highlighting the correct comparison stimulus in each trial. (In pre-training the stimuli were geometric shapes and cardinal numbers. Stimuli were Kanji symbols and English letters in the training condition.) Feedback was delivered after each of these instruction trials. Next, participants practiced in a 20-s drill with 0.1-s ITI's (if they were in R+Ac condition) or in a drill with 6 consecutive trials with 3-s ITI's (if they were in Ac condition).

Immediately after a practice drill, 3s of feedback displayed the number of correct and incorrect responses in that drill. If an error occurred during the practice drill, that sample stimulus was presented in an instruction exercise that immediately followed feedback. The number of instruction trials was not always equal to the number of errors in each drill because one instruction trial was presented for each stimulus that occasioned an inaccurate response no matter how many total errors occurred with respect to that stimulus in the preceding drill. For example, if a participant made 5 errors during a drill but 4 were in the presence of the same stimulus, only 2 instruction trials would be delivered after that drill. Multiple errors on the same stimulus were not possible in the Ac condition because there were only 6 trials, each with a different stimulus, presented within a drill.

Throughout the remainder of the condition, drills continued and instruction exercises followed drills as necessary until training ended. During R+Ac conditions, training ended when participants met criteria for both rate and accuracy (14 correct and 0 incorrect) in one 20-s drill. In the yoked Ac conditions, training ended when participants completed the same number of correct practice trials and the same total number of instruction trials as in the previous R+Ac condition.

Yoking of correct practice trials was accomplished as follows. The practice component in Ac was terminated when a participant completed the same number of accurate responses during practice drills as in the previous R+Ac condition. The yoking procedure for instruction trials included delivering the same number of prompted response opportunities as in the previous R+Ac condition. All participants made fewer errors during Ac conditions relative to the R+Ac conditions to which they were yoked, which resulted in the delivery of fewer instruction trials during practice in Ac conditions. To balance this, instruction trials were delivered after the last practice drill in Ac until the participant received the same number as in the previous R+Ac condition. For example, if a participant received 10 instruction trials (6 of which were delivered at the beginning of the condition and 4 trials delivered after errors in drills) in an R+Ac condition and no errors were made during the following Ac condition, the participant (who already received 6 instruction trials at the beginning of the session) would be presented 4 more instruction trials after the last Ac practice drill.

Testing

Immediately after reaching training criteria participants were given rate, endurance, stability, and application tests. All tests used 0.1-s ITI's. The stimulus array in rate, endurance and stability test conditions was the same as in practice drills. The duration of rate and stability tests was 20s, while the duration of the endurance and application tests was 60s.

The rate and endurance tests were identical to practice drills in R+Ac, but participants did not receive performance feedback or instruction trials after the drill (and the duration of the endurance tests was 40s longer). During stability tests audible distracters in the form of the names of the English letters of the set being tested were spoken by a computer generated voice. In the application test each trial began with the computer dictating a 1 or 2-syllable word twice. The word, with one letter missing and a "_" in its place, appeared with the second dictation. The participant pressed one of three Kanji symbols presented below the word. (During pre-training, two-digit numbers were dictated and presented with one digit missing, along with three geometric shapes). See

Figure 2 for an example of a trial from an application test. When the application test was completed, the session ended.

At the beginning of all sessions after the first, participants were tested over the set from the previous session. Performance on the delayed rate test was compared with the previous rate test. If the rate of correct responses was lower on the delayed test, R+Ac training was conducted until it matched or exceeded the previous rate. After meeting the criteria, training on a new set began. If performance on the delayed rate test matched or surpassed the initial rate test, training on a new set began after the delayed application test.

CHAPTER 3

RESULTS

Of the entire experiment, all but two sessions were conducted on successive days. (For 1 participant, session 2 was conducted 6 days after the first and for another participant; the delay between sessions 2 and 3 was 2 days.) Performance of these participants after these delays was comparable to performance in sessions conducted on consecutive days. Participants performed similarly on training and tests, and the following figures display data from 2 participants collected over 5 consecutive days. Training data are shown in Figures 3 through 6. Figures 6 and 7 are comprised of test data. Data for the rest of the participants is in Appendix A.

Training Results

Figure 3 displays the number of correct responses during the practice component of training and the duration of training (from the initial six instruction trials until meeting criteria) in each session. Subject 12 completed 880 correct practice trials in the first R+Ac condition (session 1) and the same number of practice trials in the yoked Ac condition (session 2). In the second R+Ac condition (session 3), Subject 12 made 364 accurate responses during training; thus, accurate responses in the second Ac condition (session 4) were yoked to this number. The duration of the training portion of session 1 (R+Ac) was 43 min. The length of time required to meet mastery criteria in session 2 (Ac) was 85 min. Durations of subsequent R+Ac and Ac sessions were 18 and 34 min, respectively. For all participants the number of trials to mastery and duration of training in the second R+Ac condition was much less relative to the first.

No trials were yoked for Subject 3 who completed 634, 357, 114, and 128 correct practice trials across all sessions (all R+Ac). Subject 3 reached the mastery criteria in 39 min in session 1, and 19, 6, and 7 min in sessions 2, 3, and 4, respectively. Decreasing trends were apparent in the number of trials to mastery in R+Ac conditions and in duration of sessions for all subjects.

Number of errors and instruction trials are graphed in Figure 4. Subject 12 made 32 errors during training of the first R+Ac condition. During training in session 2 (Ac), Subject 12 made 10 errors. Subject 12 made 15 errors in session 3 (R+Ac) and 3 errors in session 4(Ac). In sessions 1 and 2, Subject 12 completed 27 yoked instruction trials, and 20 in sets 2 and 4. Although there was an overall decreasing trend in error rates, frequencies of errors in Ac training sessions were also lower compared to the R+Ac sessions to which they were yoked. Subject 3 made 53, 14, 1, and 3 errors across training sessions (R+Ac). Total number of instruction trials completed by Subject 3 across sessions was 44, 19, 7, and 9. Data indicate decreasing trends in error rates across repeated R+Ac sessions.

Number of correct responses and average latencies/drill are shown in Figures 5 and 6, and Tables 1 and 2. Average latencies were calculated by dividing the sum of the accurate response latencies by the number of accurate responses in each drill. To meet the mastery criteria- 14 accurate responses in 20s- average latencies would need to be at or below 1.4s. Latencies from incorrect responses were omitted from these calculations. Performance data of Subject 12 are graphed in Figure 5, and Figure 6 consists of data from Subject 3. Table 1 includes the total number of drills, ranges of averaged latencies of accurate responses, and ranges of the number of accurate responses by Subject 12 Performance data of Subject 3 are listed in Table 2. Data from the first and last 20 drills of each session are presented. Data from these drills were selected because they reveal acquisition of the conditional discriminations both by the increasing numbers of correct responses/drill as well as by the decreasing averaged latencies of responses. Exceptions include both participants meeting mastery in less than 40 drills and in less than 20 drills for Subject 3.

In the first 20 drills of Subject 12's first session (R+Ac), the number of correct responses/drill ranged from 2 to 9. In the final 20 drills of session 1, rate of accurate responses ranged from 5 to 14. Subject 12 met mastery criteria after 97 drills in session 1. The range of accurate responses in the initial 20 drills of session 2 (Ac) was 4 to 6. In the last 20 drills of session 2, Subject 12 made no errors and made six accurate responses in each drill, except for the last drill. In that 149th drill, there were only two trials presented due to the yoking procedure. In the first 20 drills of session 3 (R+Ac), Subject 12's rate of correct answers ranged from 3 to 13 (missing the mastery criteria by one accurate response in drill 16). Subject 12's performance met the mastery criteria in the 37th drill, and accurate responses ranged from 7 to 14 in those last 17 drills. In session 4 (R+Ac), Subject 12's accurate responses ranged from 5 to 6 in the first 20 drills and remained at 6 throughout the remainder of the session. Subject 12 completed training after 64 drills in session 4.

The averaged latencies of accurate responses made by Subject 12 in the first 20 drills of session 1 ranged from 4.8s to 1.9s. In the final 20 drills in the first session, average latencies ranged from 2s to 1.3s. In session 2 (Ac), averaged latencies ranged from 5.1s to 1.9s in the first 20 drills and 1.8s to 1.1s in the final 20. In the next session

(R+Ac), the range of averaged latencies was 3.7s to 1.4s in the first 20 drills, and 2.6s to 1.3s in the last 17 drills. In session 4 ranges of averaged latencies in the first and last 20 drills were 3.6s to 1.3s and 1.7s to 1.2s.

Ranges of accurate responses of Subject 3 were 3 to 9 in the first 20 drills of session 1, and 6 to 14 in the last 20 of 71 drills. In session 2, accurate responses in the first 20 drills ranged from 5 to 12 and 7 to 14 in the last 18 of a total of 38 drills. Rate of accurate responses ranged from 5 to 14 in the 12 drills of session 3 as well as the 13 drills of session 4.

Averaged latencies of Subject 3's accurate responses ranged from 3.7s to 1.7s in the first 20 drills of session 1 and from 2.9s to 1.3s in the last 20 drills. In the first 20 drills of session 2 averaged latencies ranged from 3.6s to 1.5s, and 2.5 to 1.3s in the last 18 drills. Averaged latencies ranged from 3.6s to 1.3s in session 3 and 2.9s to 1.3s in session 4.

Testing Results

Test data are displayed in Figures 7 and 8. The duration of rate and stability tests was 20s, and performance data from those tests were multiplied by 3 to yield count/minute. Figure 7 contains data from tests conducted immediately after training. Performance data collected during delayed tests are graphed in Figure 8. It should be noted that slight delays occurred during stability tests. When a participant responded while the computer generated voice was saying a letter name, the pronunciation of the letter name was completed before moving to the next trial. This generated an artificial ceiling and resulted in overall lower response rates in the stability test conditions.

The following data are displayed in Figure 7. On rate tests after training in session 1 (R+Ac), Subject 12 made 42 accurate responses/min and zero errors. Performance in session 2 (Ac) was 45 correct responses/min with zero errors. After training in session 3 (R+Ac), Subject 12 made 42 correct responses/min and zero errors. In session 4 (Ac), Subject 12 scored 48 accurate responses/min and zero errors. Rates of Subject 12’s accurate responses on endurance tests across sessions were 43, 44, 38, and 39/min, again with no errors. Stability test results include 27 correct responses and 3 errors/min after training in session 1 (R+Ac), and 33 accurate responses with zero errors in remaining conditions. On application tests Subject 12 made 15 and 17 accurate responses with no errors in sessions 1 and 3 (both after R+Ac training), and 14 accurate responses and one error in sessions 2 and 4 (after Ac training).

Subject 3’s rates of accurate responses on rate tests conducted immediately after training across sessions (all R+Ac) were 39, 33, 45, and 33/minute. There were no errors on rate tests. Correct response rates on endurance tests were 38, 34, 38, and 35. Subject 3 made one inaccurate response on the endurance test after training in session 2, and 2 errors after session 3 training. On stability tests, there were 24 accurate responses and 6 errors/min after training in session 1, and 30 accurate responses/min with 0 errors in sessions 2, 3, and 4. Rates of correct responses on application tests were 16 after training in sessions 1 and 4 and 15 correct responses/min after training in sessions 2 and 3. There were no errors on application tests.

Data from delayed tests are presented in Figure 8. Rate of correct responses on rate tests conducted one day after Subject 12 completed training in sessions 1, 2, and 4 was 42, and 45 correct responses/min 1 day after session 3. There were no errors on delayed rate or stability tests. One day after training sessions, endurance test performance was 43, 46, 45, and 42 accurate responses/min, with one error in the third and fourth tests. Rate of responses were 17 correct/min with 0 errors on application tests given one day after sessions 1 and 3, 18 correct and 1 incorrect/min for application tests given 1 day after session 2, and 14 correct and 1 incorrect/min 1 day after session 4.

Rate of accurate responses by Subject 3 on delayed rate tests were 39 after sessions 1 and 2, and 36 and 33 after sessions 3 and 4. There were 3 errors/min on delayed rate tests after sessions 1 and 4 and 0 errors after sets 2 and 3. Rate of correct responses on delayed endurance tests per session was 40, 38, 43, and 37/min. Subject 3 made one error on the endurance test 1 day after session 2, and no other errors on delayed endurance tests. On delayed stability tests, there were 27 correct responses/min after sessions 1 and 3, 30 after session 2, and 33 after session 4. The only occurrence of errors in these tests was 3 errors/min after session 3. Rate of accurate responses on delayed application tests was 16/min after sessions 1, 2, and 3, and 17/min after session 4. There was one error in the test after session 2. Figure 9 displays performance of all subjects across all tests and sessions..

CHAPTER 4

DISCUSSION

Decreasing trends were evident in number of trials to mastery and frequency of errors across all sessions, as well as in durations of R+Ac sessions. Overall decreasing trends could have been the result of sequence or practice effects. Performance data from Subject 3 displays a decreasing trend until session 4, although the number of trials to mastery, frequency of errors, and durations of sessions 3 and 4 were similar. Although these trends are generally decreasing, there is also some evidence that the training condition made a difference. Regarding errors, for participants who received yoked trials, the frequency of errors in each Ac condition was lower relative to that in the previous R+Ac condition.

Another difference between R+Ac and Ac conditions was the durations of those sessions. The durations of sessions in Ac conditions were more than twice that of the R+Ac conditions to which they were yoked. Thus, R+Ac conditions were more efficient in that less time was required to produce similar outcomes. Learning a new skill typically involves frequent practice attempts. When comparing teaching methods that produce similar outomes, “from a technology standpoint… the less time-consuming procedure (i.e., freely emitted responses) is more efficient and therefore practically a superior procedure” (Binder, 2004, p. 285).

Although latencies of responses by Subject 12 decreased during drills in R+Ac conditions, latencies also decreased in Ac conditions. It is possible that initiating experiments with R+Ac conditions resulted in shorter latencies in subsequent Ac conditions. By establishing a history of reinforcement for high rates of responding, it is possible that performance of new similar skills would resemble what has been reinforced, in this case, short latencies. It is possible that test conditions functioned in this way, as discussed below. Initial and successive experimental conditions without rate criteria would be required to address this possible sequence effect. Regardless, training in Ac conditions produced equally high rates of responding in the test conditions and equally low latencies during training as in the R+Ac conditions throughout this experiment.

Repeated exposure to tests provide opportunities to practice in varied conditions. As the amount of practice in these conditions increases, it is likely that the quality of performance will improve.

Performance on tests was similar across all participants, regardless of the inclusion of rate in the mastery criteria. Figure 9 displays the similarities of test performance across participants and lack of differentiated results of training conditions.

Repeated exposure to test conditions may have affected performance in a number of ways by establishing histories of reinforcement of high-rate responding in those conditions. These exposures may result in increases in the likelihood of responding at high rates for extended durations of time or performance that is less sensitive to distraction. The latter is especially likely in the stability test condition. Rate of accurate responses by all participants was considerably lower in the first stability test compared to all other attempts. These increasing trends are illustrated in Figure 10. Performance improved across repeated exposure to the stability test conditions. This suggests a decrease in the probability that the computer generated voice interrupted performance. Since tests can be considered as a form of practice, sequence effects of test conditions must be considered in this area of research.

There was also little difference between performance on tests conducted immediately after training and delayed tests. Tests conducted more than 1 day after training may indicate that the length of the delays in this experiment was not sufficient to produce differentiated performance.

These data do not support that measurably superior outcomes are produced when rate is included in a mastery criterion. It is possible that the small number of stimuli in each set made acquisition and mastery more likely, no matter the specific criteria. Also, the relatively short drill durations and relatively low terminal rate (42/min) may not have been adequate to produce differentiated performance across experimental conditions. However, raising the rate criterion would have increased the number of trials to mastery, possibly significantly. In this experiment, practicing the same number of accurate responses under different training conditions produced similar outcomes.

There may be several different training procedures that result in fluent performance, but they all have a commonality, namely practice. Several, dozens, or hundreds of practice opportunities may be required for a skill to be considered mastered. It is possible that fluent performance results from a sufficient amount of practice, whether or not rate-building procedures are used.

Figure 1. An example of the stimulus arrangement. The space in the lower right corner is due to the capacity of the program to display 7 comparison stimuli. This placement was constant throughout training and tests, except for the application test.

Figure 2. Example of application test trial.

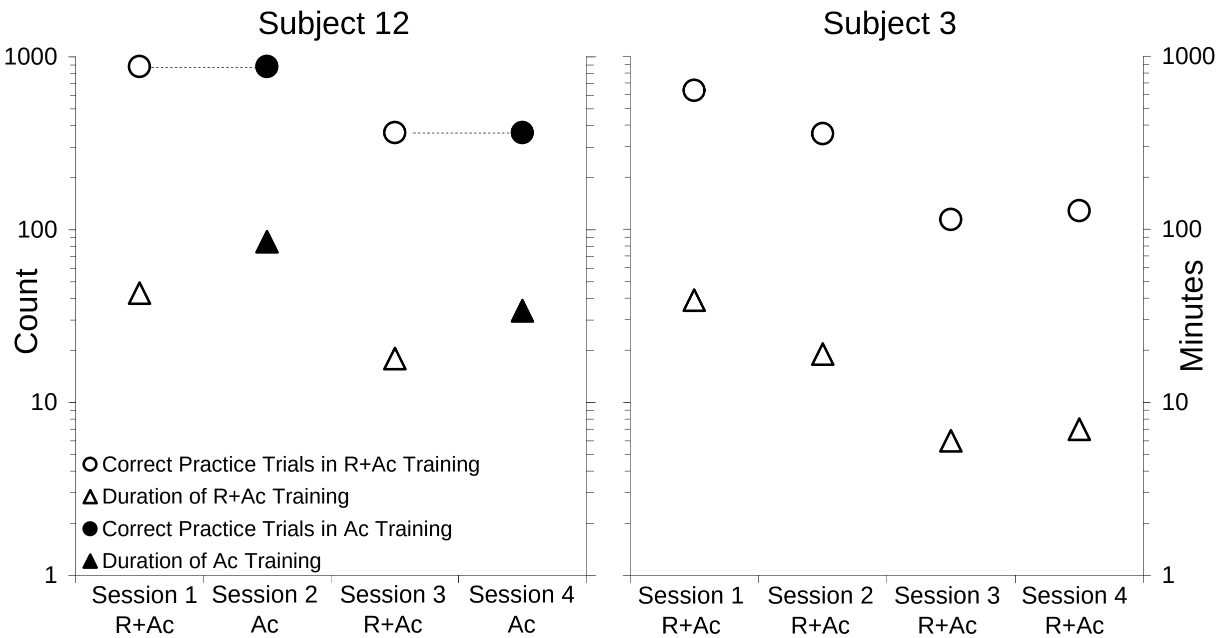

Figure 3. Total number of correct practice trials per session and duration of sessions for experimental subject (Subject 12) and control subject (Subject 3). The circles indicate the number of correct practice trials and the triangles the duration of training. For Subject 12, the number of correct practice trials in Ac was yoked to R+Ac, as shown by the dashed lined.

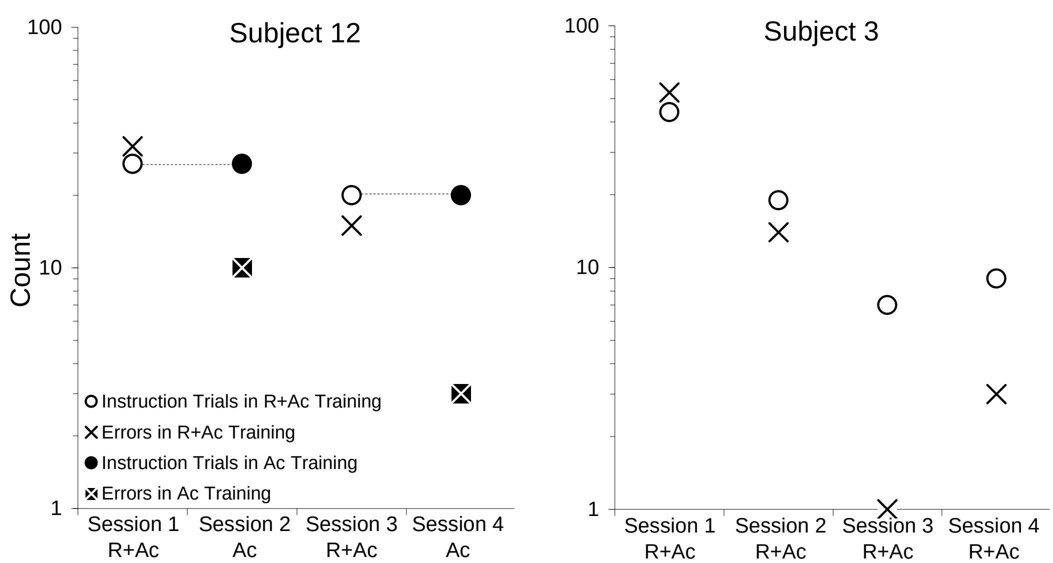

Figure 4. Number of errors and instruction trials from training sessions of experimental subject (Subject 12) and control subject (Subject 3). The dotted line connects the yoked number of instruction trials. In the first session for both participants, frequency of errors was higher than training trials due to multiple errors on a sample stimulus in the same drill.

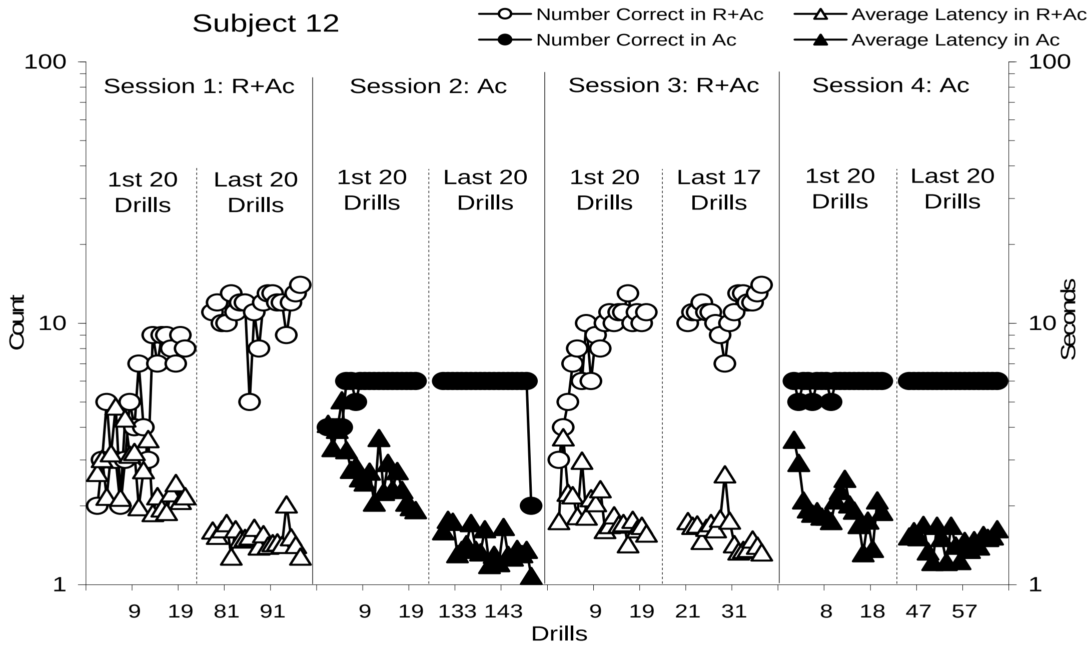

Figure 5. Number of correct responses per drill and averaged latencies of correct responses during drills across training sessions for Subject 12. Circles represent the number of correct responses. Triangles indicate the averaged latencies of all correct responses per drill. Subject 12 met mastery criteria in 37 drills in the third session.

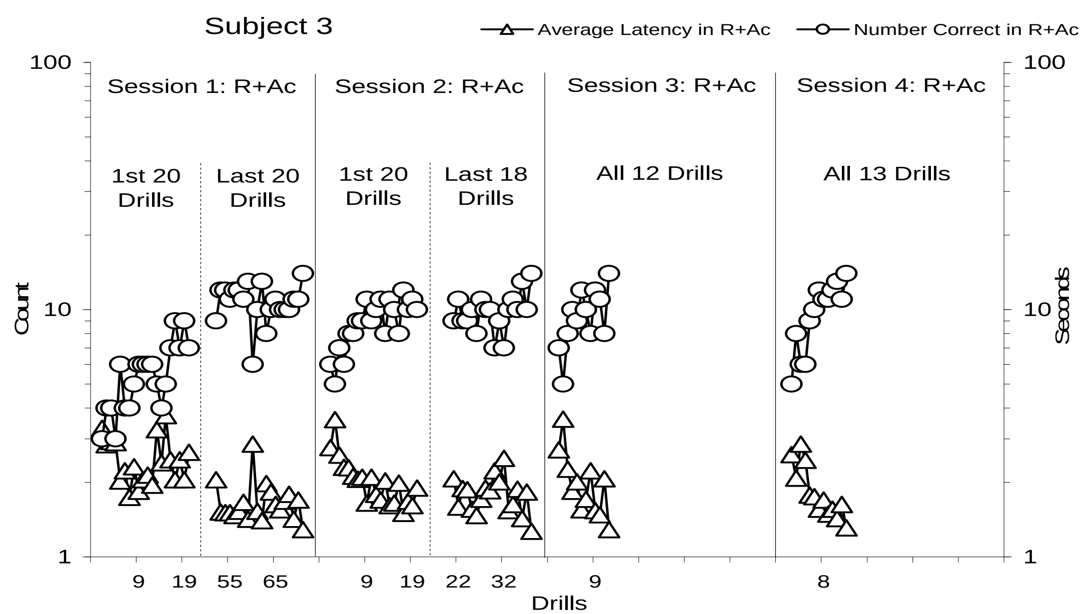

Figure 6. Number of correct responses per drill and averaged latencies of correct responses of drills across sessions for Subject 3. Circles reflect the number of correct responses. Triangles represent the averaged latencies of all correct responses per drill. Subject 3 met mastery criteria in 12 drills in the third session and 13 drills in the fourth session.

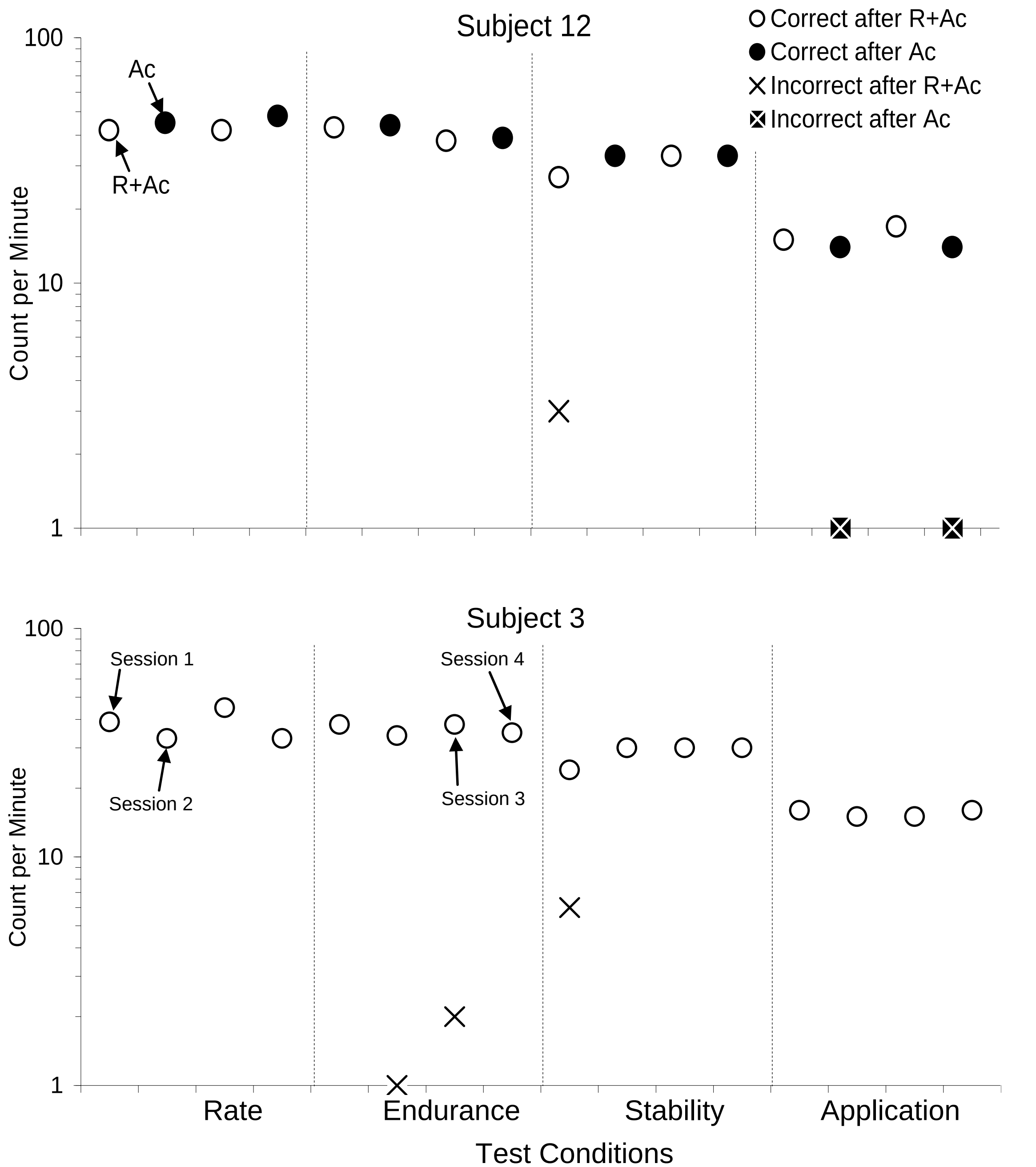

Figure 7. Rate of correct responses and errors for each test conducted immediately after meeting mastery criteria for Subject 12 (upper graph) and Subject 3 (lower graph). Open circles indicate the rate of correct answers after R+Ac Training. Correct responses on tests after Ac training are represented by closed circles.

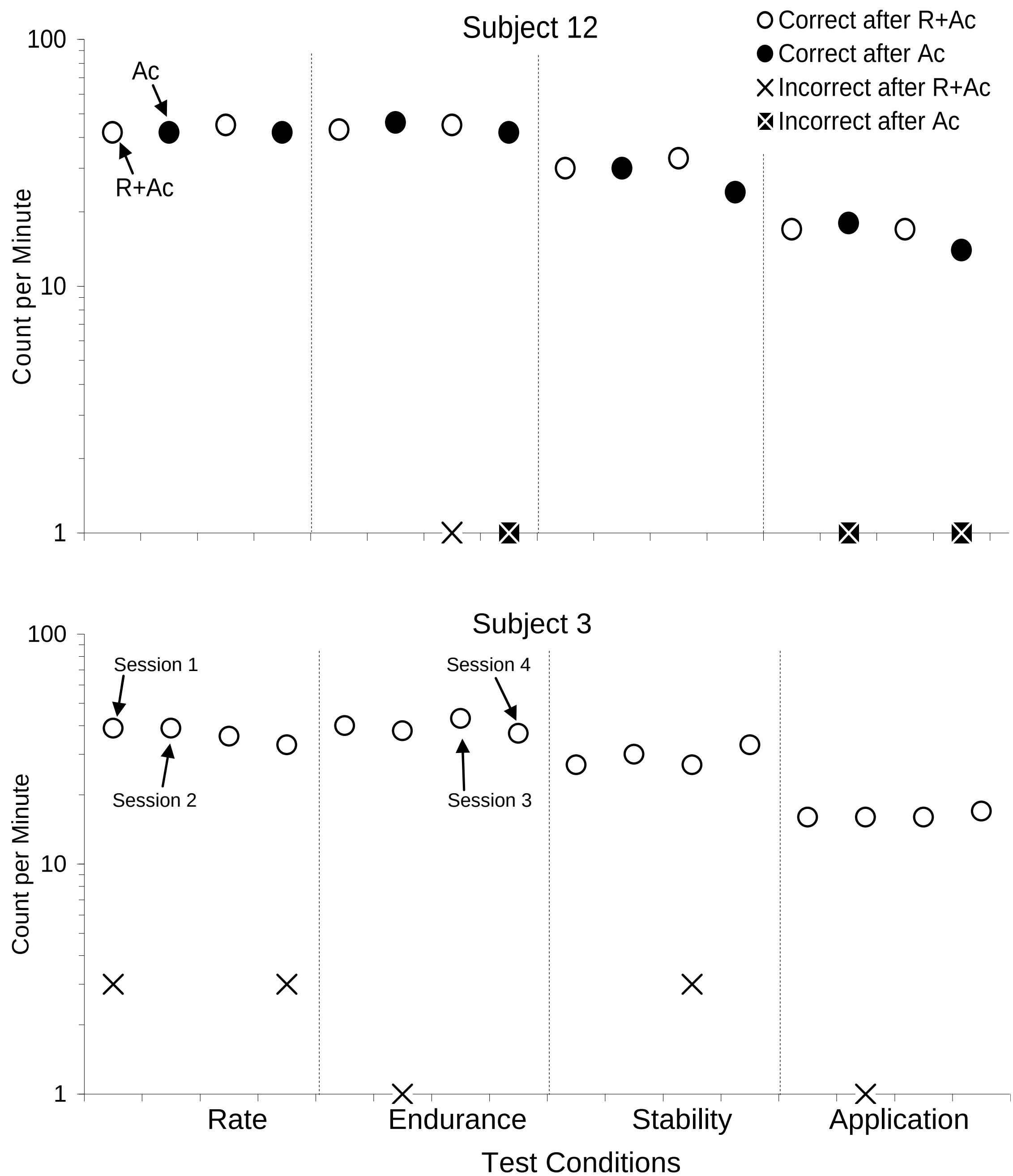

Figure 8. Rate of correct responses and errors on delayed tests. The rate of correct responses after R+Ac training are represented by open circles. Closed circles refer to the rate of correct answers on delayed tests after Ac training.

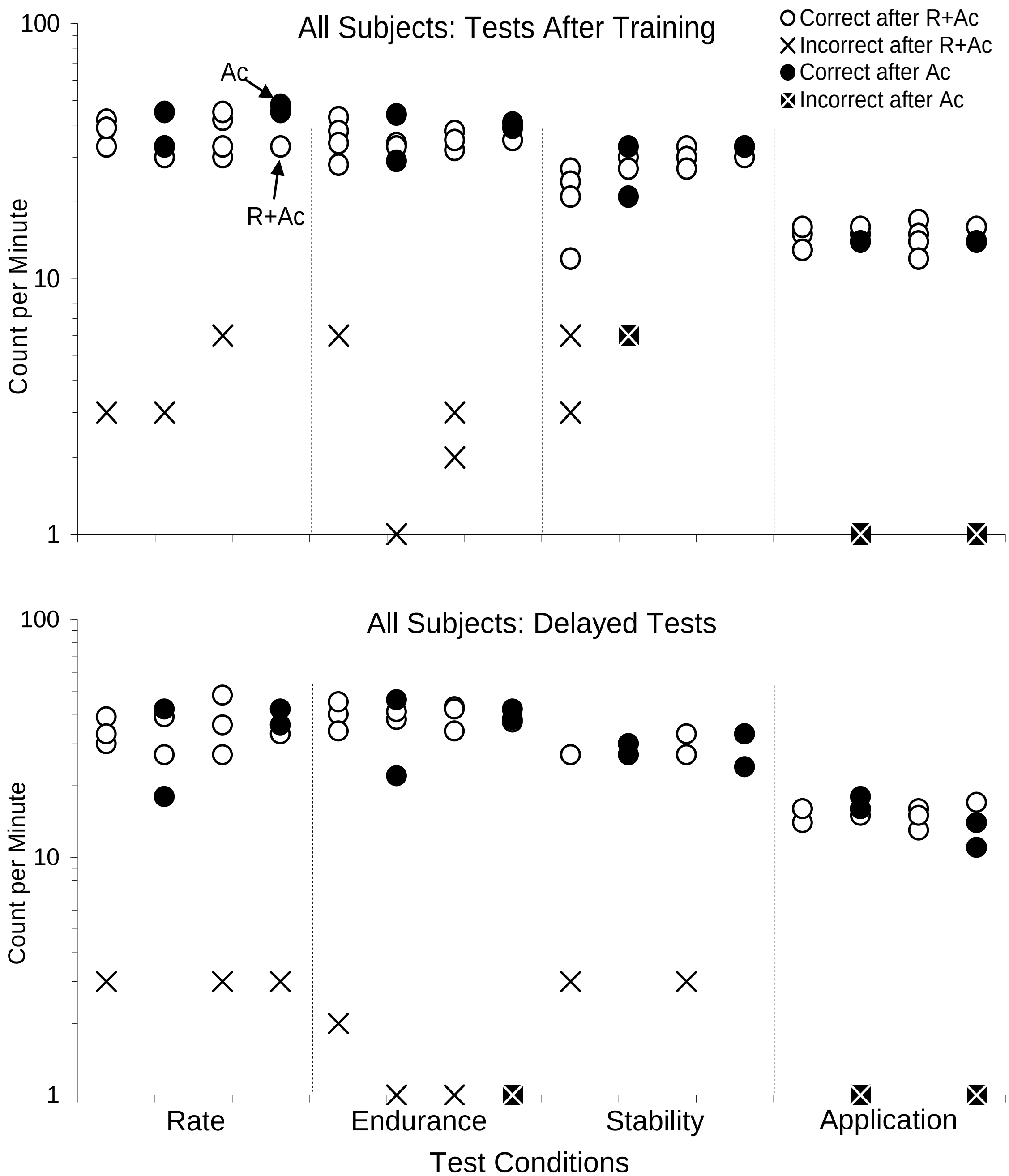

Figure 9. Rate of all participants' correct responses and errors on all tests. The rate of correct responses after R+Ac training are represented by open circles. Closed circles indicate the rate of correct answers on tests after Ac training.

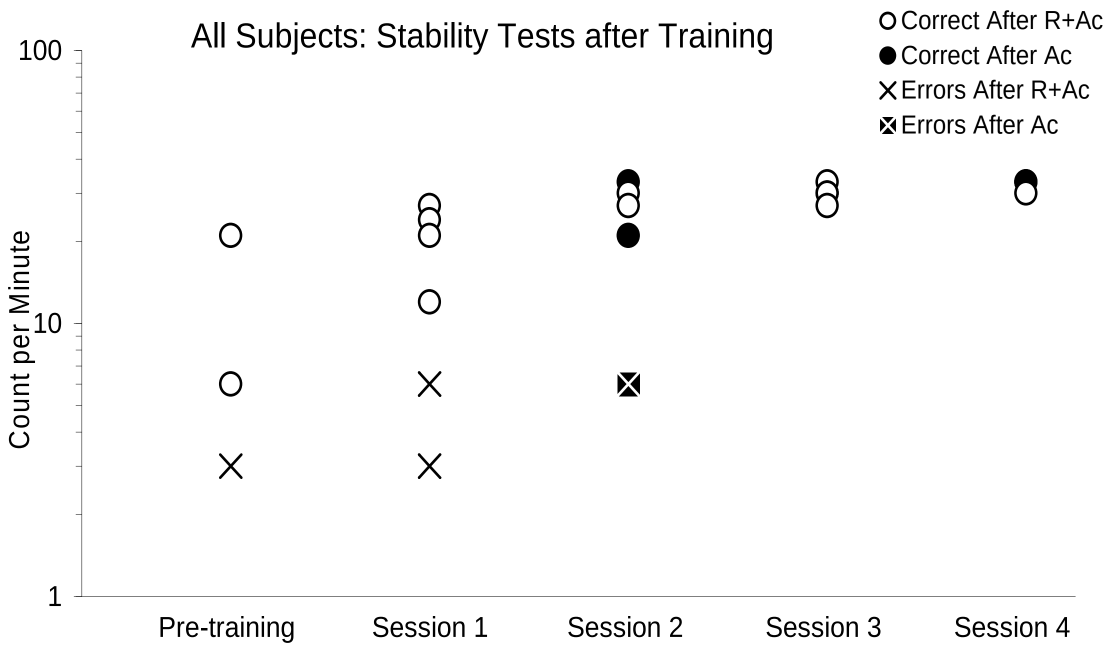

Figure 10. Performance of all participants in stability test conditions conducted after training. Rate of correct responses in R+Ac conditions are represented by open circles, and closed circles refer to the rate of correct responses after Ac training.

Table 1

Training Performance of Subject 12

Subject 12		Range of Averaged Latencies	Range of Accurate Responses	Total Number of Drills
Session 1 (R+Ac)	1st 20 Drills	4.8s - 1.9s	2-9	97
	Last 20 Drills	2s - 1.3s	5-14	
Session 2 (Ac)	1st 20 Drills	5.1s - 1.9s	4-6	149
	Last 20 Drills	1.8s - 1.1s	2-6	
Session 3 (R+Ac)	1st 20 Drills	3.7s - 1.4s	3-13	37
	Last 17 Drills	2.6s - 1.3s	7-14	
Session 4 (Ac)	1st 20 Drills	3.6s - 1.3s	5-6	64
	Last 20 Drills	1.7s - 1.2s	6-6	

Note. Only latencies of accurate responses are included in the averages.

Table 2

Training Performance of Subject 3

Subject 3		Range of Averaged Latencies	Range of Accurate Responses	Total Number of Drills
Session 1 (R+Ac)	1st 20 Drills	3.7s - 1.7s	3 - 9	71
	Last 20 Drills	2.9s - 1.3s	6 - 14	
Session 2 (R+Ac)	1st 20 Drills	3.6s - 1.5s	5-12	38
	Last 18 Drills	2.5s - 1.3s	7 - 14	
Session 3 (R+Ac)	All 12 Drills	3.6s-1.3s	5-14	12
Session 4 (R+Ac)	All 13 Drills	2.9s-1.3s	5-14	13

Note. Latencies of inaccurate responses are not included in the averages.

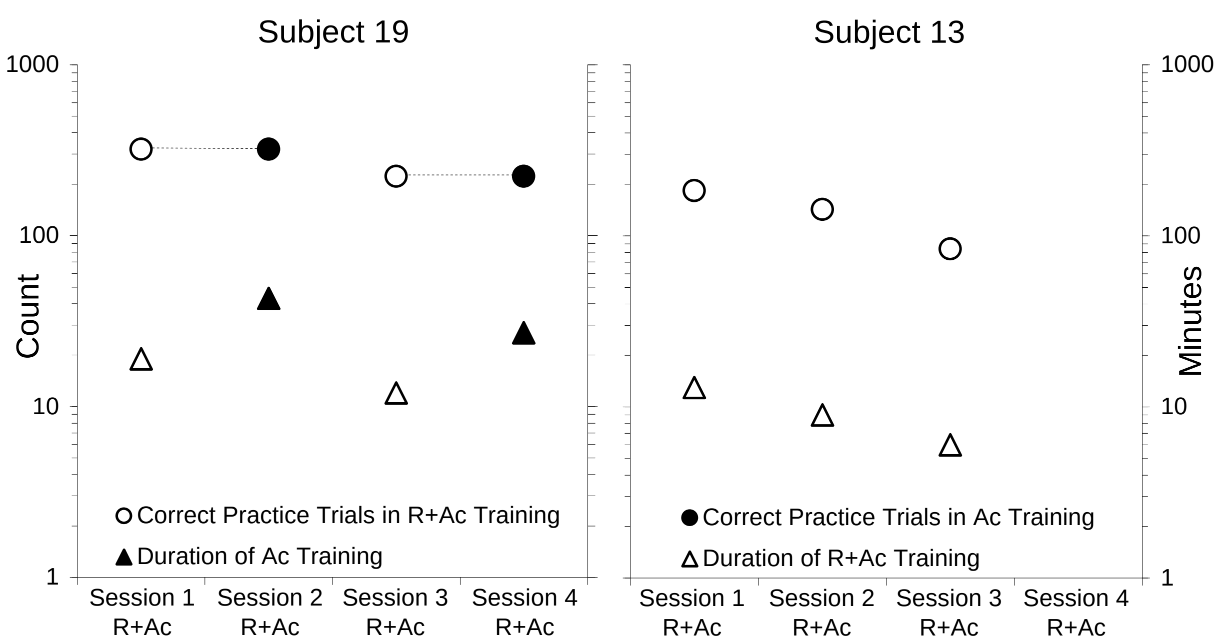

Figure 11. Total number of correct practice trials per session and duration of sessions for experimental subject (Subject 19) and control subject (Subject 13). The circles indicate the number of correct practice trials and the triangles the duration of training. For Subject 19, the number of correct practice trials in Ac was yoked to R+Ac, as shown by the dashed lined.

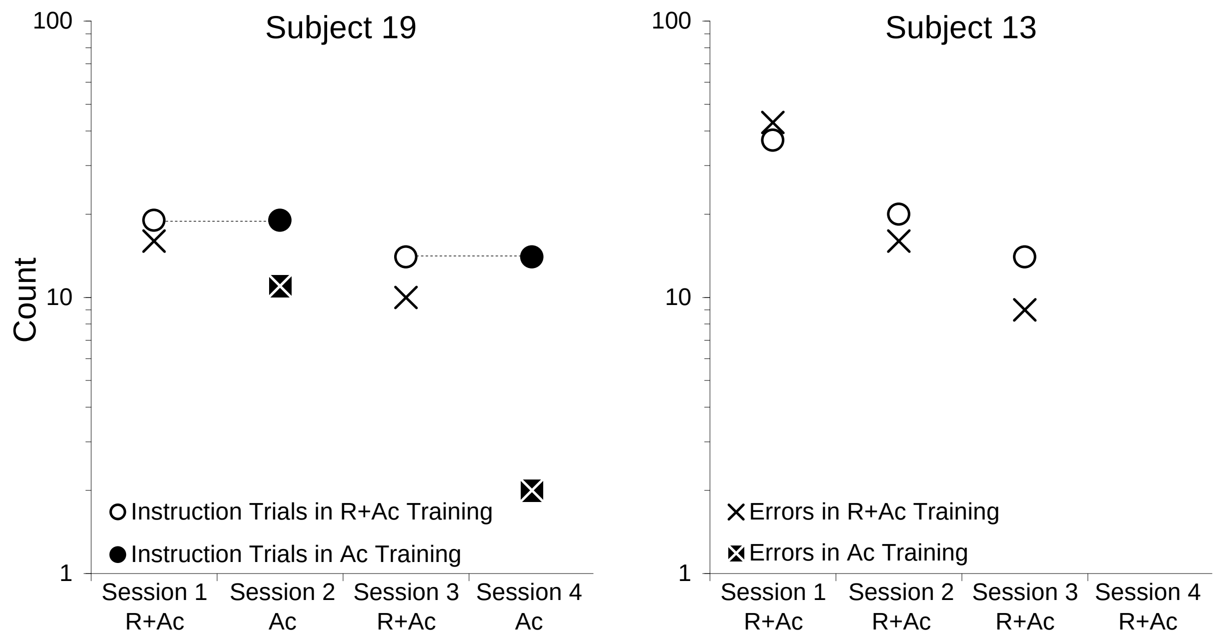

Figure 12. Number of errors and instruction trials from training sessions of experimental subject (Subject 19) and control subject (Subject 13). The dotted line connects the yoked number of instruction trials. In the first session for both participants, frequency of errors was higher than training trials due to multiple errors on a sample stimulus in the same drill.

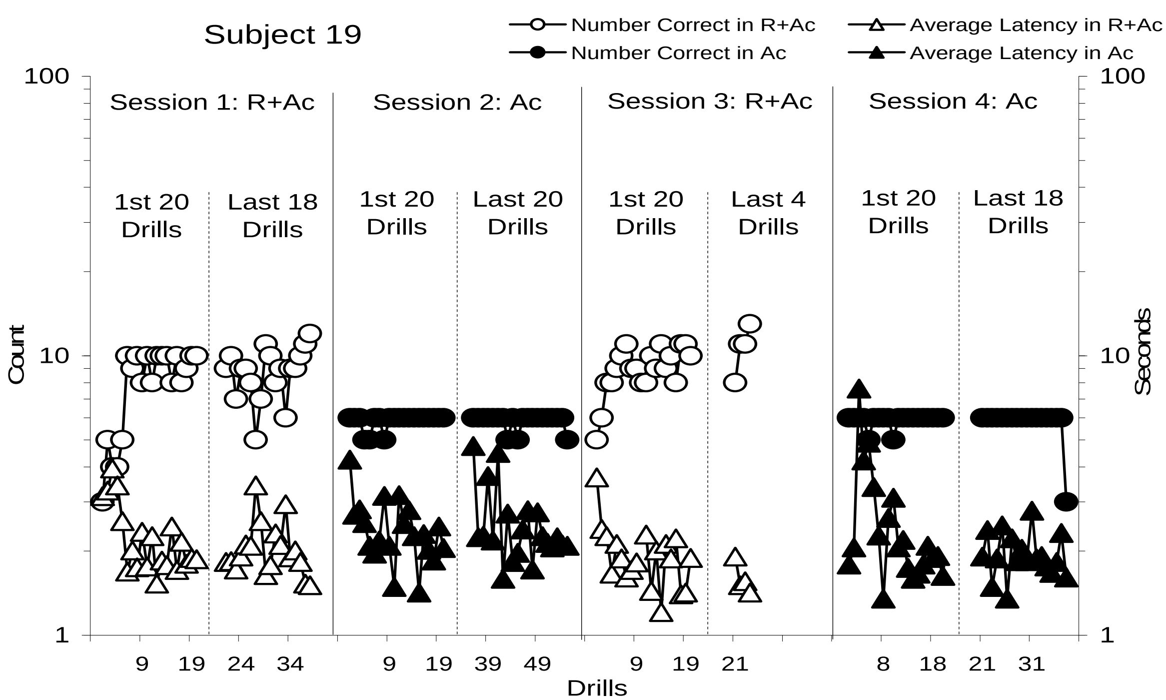

Figure 13. Number of correct responses per drill and averaged latencies of correct responses of drills across sessions for Subject 19. Circles reflect the number of correct responses. Triangles represent the averaged latencies of all correct responses per drill.

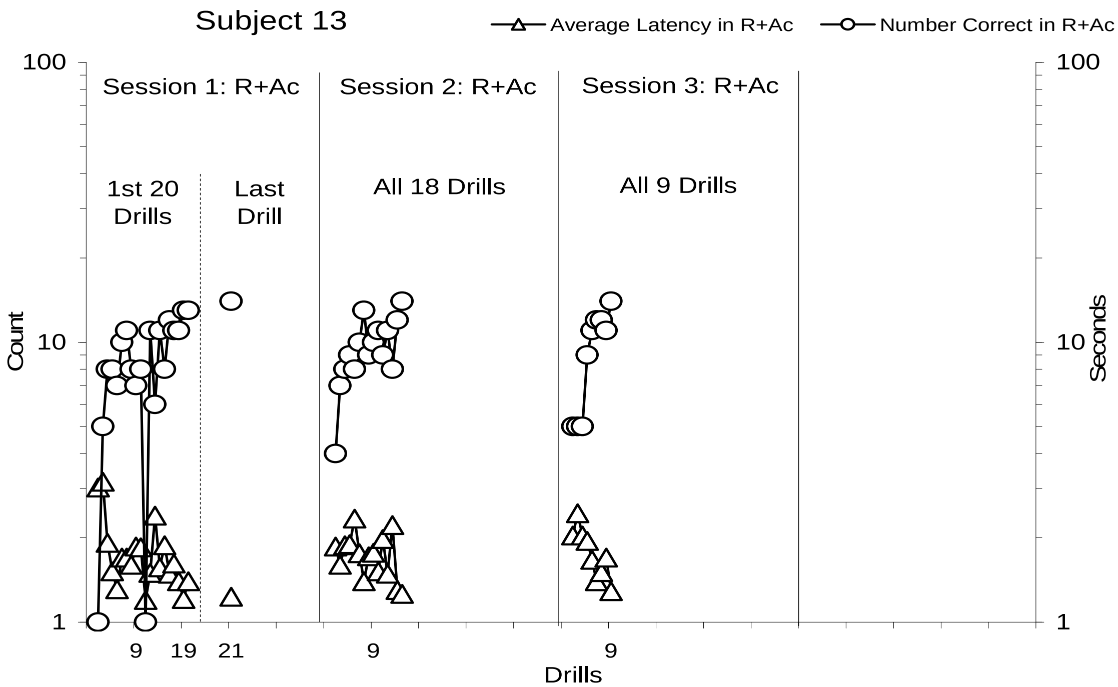

Figure 14. Number of correct responses per drill and averaged latencies of correct responses of drills across sessions for Subject 13. Circles reflect the number of correct responses. Triangles represent the averaged latencies of all correct responses per drill.

Table 3

Training Performance of Subject 19

Subject 19		Range of Averaged Latencies	Range of Accurate Responses	Total Number of Drills
Session 1 (R+Ac)	1st 20 Drills	4s - 1.5s	3 - 10	38
	Last 18 Drills	3.4s - 1.5s	5 - 12	
Session 2 (Ac)	1st 20 Drills	4.3s - 1.4s	5 - 6	55
	Last 20 Drills	4.8s - 1.6s	5 - 6	
Session 3 (R+Ac)	1st 20 Drills	3.7s - 1.2s	5 - 11	24
	Last 4 Drills	1.9s - 1.4s	8 - 13	
Session 4 (Ac)	1st 20 Drills	7.6s - 1.4s	5 - 6	38
	Last 18 Drills	2.8s - 1.4s	3 - 6	

Note. Latencies of inaccurate responses are not included in the averages.

Table 4

Training Performance of Subject 13

Subject 13		Range of Averaged Latencies	Range of Accurate Responses	Total Number of Drills
Session 1 (R+Ac)	All 21 Drills	3.2s - 1.2s	1 - 13	21
Session 2 (R+Ac)	All 15 Drills	2.3s - 1.3s	4 - 14	15
Session 3 (R+Ac)	All 9 Drills	2.5s - 1.3s	5 - 14	9

Note. Latencies of inaccurate responses are not included in the averages.

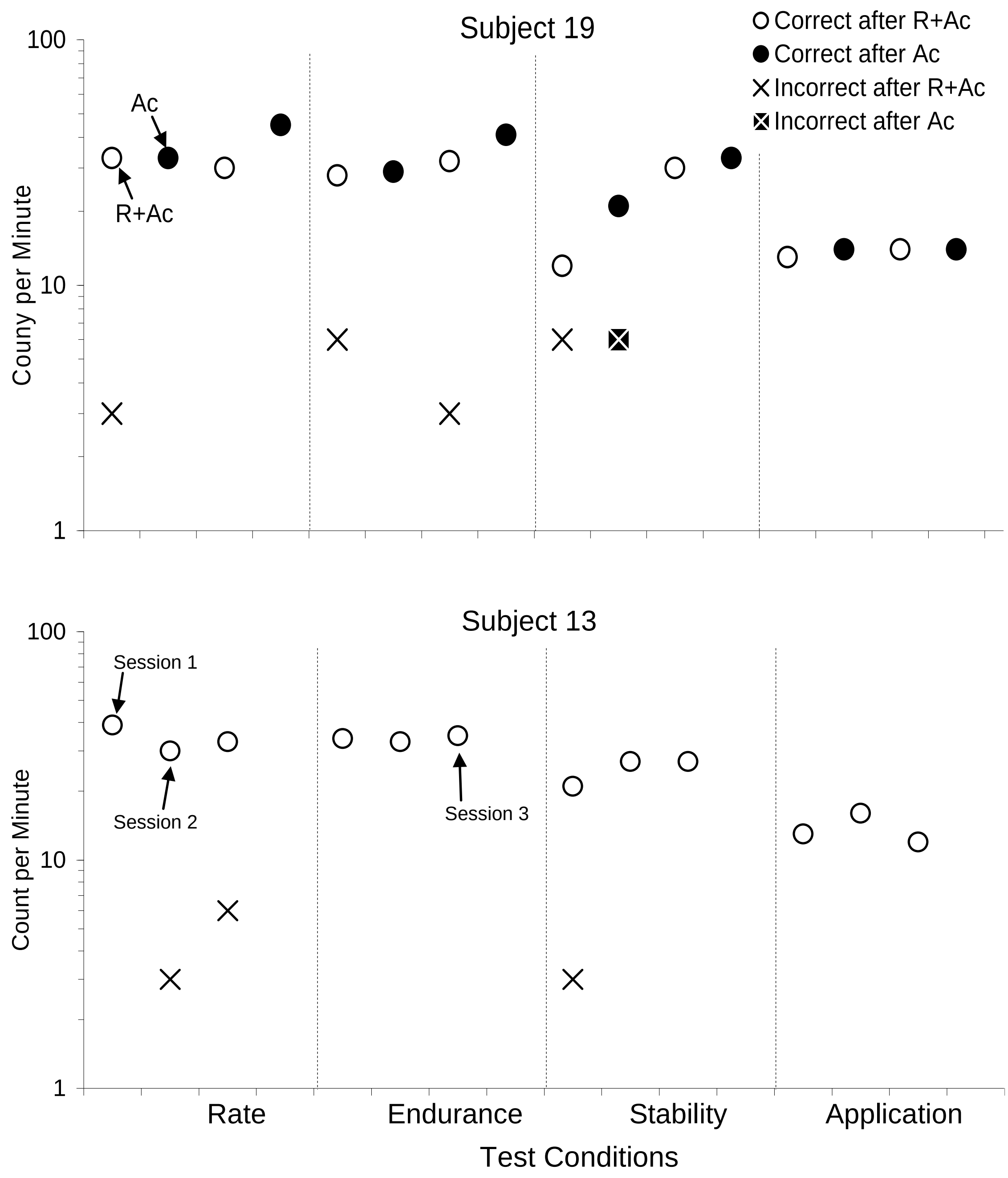

Figure 15. Rate of correct responses and errors for each test conducted immediately after meeting mastery criteria for Subject 19 (upper graph) and Subject 13 (lower graph). Open circles indicate the rate of correct answers after R+Ac Training. Correct responses on tests after Ac training are represented by closed circles.

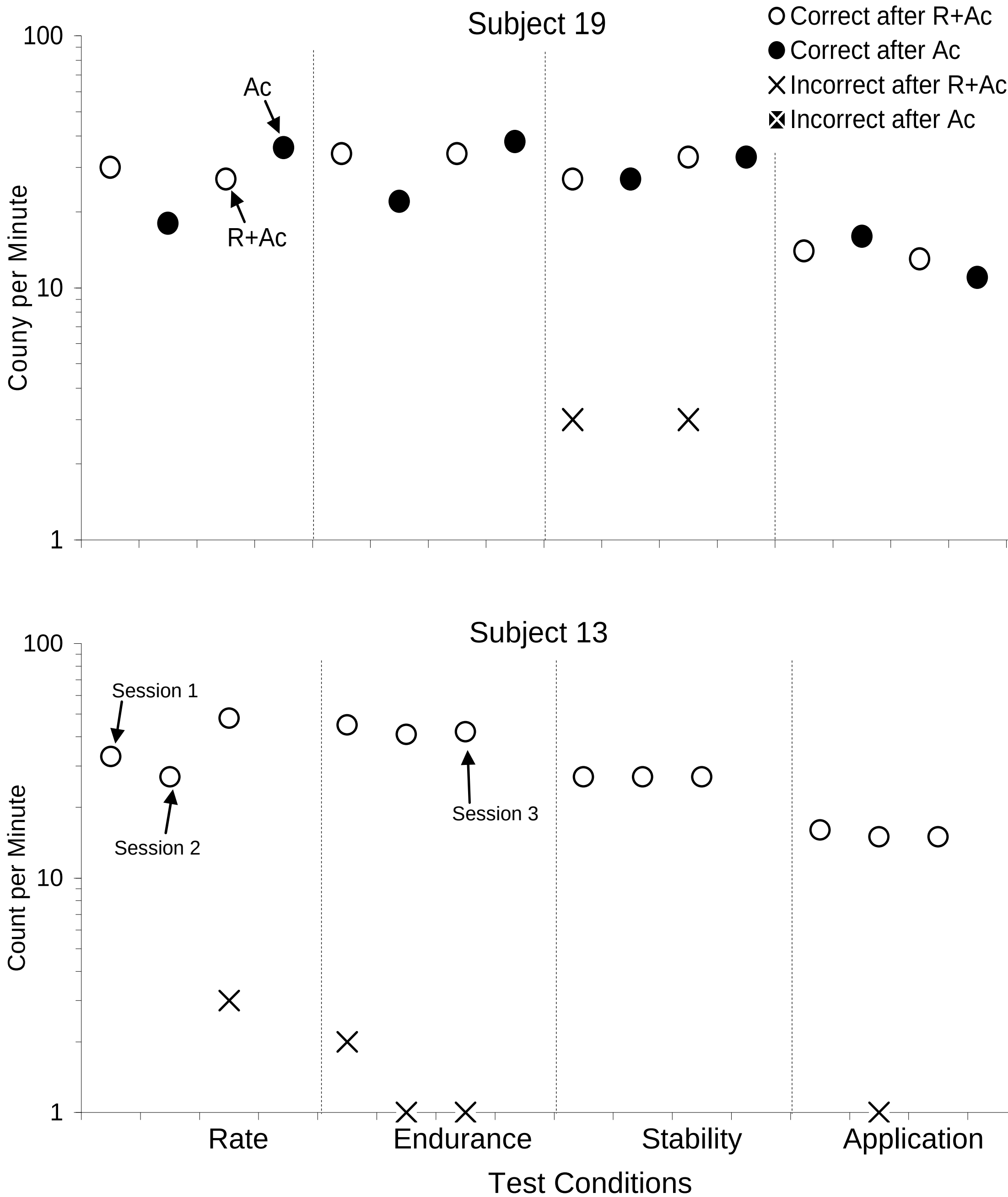

Figure 16. Rate of correct responses and errors on delayed tests by Subject 13 (lower graph) and Subject 19 (upper graph). The rate of correct responses after R+Ac training are represented by open circles. Closed circles refer to the rate of correct answers on delayed tests after Ac training.

REFERENCES

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Barrett, B.H. (2002). *The technology of teaching revisited: A reader's companion to B. F. Skinner's book*. Concord, MA: Cambridge Center for Behavioral Studies.

Binder, C. (1988). Precision teaching: Measuring and attaining exemplary academic achievement. *Youth Policy, 10*, 12-15.

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

Binder, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields (2004). *Behavior Analyst, 27*, 281-286.

Doughty, S.S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *Behavior Analyst, 27*, 7-23.

Evans, S. S., & Evans, W. H. (1985) Frequencies that ensure skill competency. *Journal of Precision Teaching, 6*, 25-30.

Evans, S. S., Merger, C. D., & Evans, W. H. (1983). The relationship of frequency to subsequent skill acquisition. *Journal of Precision Teaching, 4*, 28-34.

Haughton, E. C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching, 1*, 3-20.

Johnson, K. R. (2007, November). Frequency-based curriculum discoveries. In C. Binder (Chair), *My most interesting curriculum exploration.* Panel discussion conducted at the Twentieth International Precision Teaching Conference, Kansas City, MO.

Johnson, K. R.& Layng, T. V. J. (1996). On terms and procedures: Fluency. *Behavior Analyst*, *19*, 281-288

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

Johnson, K. R., & Layng, T. V. J. (1994). The Morningside model of generative instruction. In R. Gardner III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 173-197). Monterey, CA: Brooks/Cole.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2^{nd} ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Kubina, R. M. (2005). The relations among fluency, rate building, and practice: A response to Doughty, Chase, and O'Shields(2004). *Behavior Analyst*, *28*, 73-76.

Shirley, M. J., & Pennypacker, H. S. (1994). The effects of performance criteria on learning and retention of spelling words. *Journal of Precision Teaching, 12*, 73-86.

Wheetley, B. (2005). T*he effects of rate of response on retention, endurance, stability, and application performance on a match-to-sample task*. Unpublished master's thesis, University of North Texas, Denton.