THE EFFECTS OF COPYING BEFORE, COPYING AFTER, AND GUESSING ON ACQUISITION RATE AND RETENTION

Sarah Ellen Pinkelman, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2010

APPROVED:

Jesús Rosales-Ruiz, Major Professor
Shahla Ala'i-Rosales, Committee Member
Traci Cihon, Committee Member
Richard G. Smith, Chair of the Department of Behavior Analysis
Thomas L. Evenson, Dean of the College of Public Affairs and Community Service
James D. Meernik, Acting Dean of the Robert B. Toulouse School of Graduate Studies

Pinkelman, Sarah Ellen. The Effects of Copying Before, Copying After, and Guessing on Acquisition Rate and Retention. Master of Science (Behavior Analysis), December 2010, 50 pp., 15 figures, 2 tables, references, 12 titles.

Computer-based instructional programs are being used more frequently in classrooms. While these programs offer many benefits from traditional teaching methods, humans still need to program them. There is inconsistency in the literature regarding the best way to design such programs. The purpose of this study was to evaluate the effects of three training procedures in teaching individuals to type a specified three-letter response in the presence of a corresponding symbol. Results show that the training format that prompted individuals to copy the correct response before the opportunity to respond was more efficient than viewing the correct response after an error, or copying the correct response after an error. A discussion of the results as well as implications for classroom use is also provided.

ACKNOWLEDGEMENTS

First, I would like to thank my family for their unwavering encouragement and support throughout this project, and in all my life endeavors. I will be forever grateful to Taylor Carmer, who has been instrumental to my success and general well being throughout the past five years by providing unconditional love, encouragement, and understanding. I owe much gratitude to my advisors, Dr. Jesús Rosales-Ruiz and Dr. Shahla Ala'i-Rosales. Their passion for the field of behavior analysis has inspired me numerous ways. Dr. Rosales-Ruiz's compelling course on stimulus control was the turning point in my academic and professional career. Working with him on this project has further enhanced my interest in matters of stimulus control, instructional design and curriculum development. Dr. Ala'i-Rosales was incredibly supportive throughout my three years at UNT. She was keen in identifying my strengths and interests and arranged several opportunities that enabled me to gain priceless mentorship in teaching in higher education. They, along with Dr. Sigrid S. Glenn have been my most influential teachers. I will carry skills learned from them throughout my career. I would also like to thank Dr. Linda K. Haymes and Dr. Ronnie Detrich, who were my first mentors in the field of behavior analysis. Their guidance in the early stages of my career was invaluable. Finally, I would like to recognize my fellow students in the Department of Behavior Analysis at the University of North Texas. Specifically, I would like to mention Jonathan Friedel, who generously offered to write the computer program for this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

We are quickly approaching an era where computer-based instruction will play a prominent role in education. Computer-based instructional programs are being used with increasing frequency in classrooms (Azevedo, 2005).

Although computers can greatly improve instructional effectiveness, they still need to be programmed. The vast array of computer-based instructional programs differs dramatically on how contingencies for learning are arranged. Such inconsistencies are likely the result of disagreement among researchers regarding even the most basic instructional design components.

Perhaps the most notorious point of controversy is feedback. While many agree that feedback is important, disagreement persists concerning the most advantageous form and timing of feedback. Feedback can take on many forms, from simple praise (e.g., “Good.”) to complex error correction (e.g., “Last time you divided 100 into 20. This time divide 20 into 100.”). One point of disagreement in regard to the form of feedback is whether or not students should practice the correct response. Barbetta, Heron, and Heward (1993) designed an experiment to investigate the effects of practicing the correct response following an error. In this study, students with developmental disabilities were given index cards with words printed on them, and were asked to read the words aloud. If the student read the word correctly, praise was provided. If the student read the word incorrectly, one of two correction procedures was implemented. In one procedure, the experimenter read the word correctly, and then prompted the student to repeat the word. In the other procedure, the experimenter read the word aloud, but did not prompt the student to repeat the word. The data indicated that the

procedure that prompted students to practice the correct response produced better performance than the other procedure. Similar results have been obtained by Drevno, Kimball, Possi, Heward, Gardner, and Barbetta (1994). These researchers replicated Barbetta et al. (1993) and taught academically at-risk students definitions of science terms. The data depicted larger gains in the condition that prompted students to produce the correct response after the model. In addition, Geller (1996) also found that copying the correct response produced more accurate responding than simply displaying the correct response.

Although the above authors found that practicing the correct response was useful, others have concluded similar arrangements to be unsatisfactory. Anderson, Kulhavy, and Andre (1971) examined the effects of two training conditions on the accuracy of fill-in-the-blank and multiple-choice questions. One condition displayed the correct answer following an incorrect response. The other condition displayed the correct answer (before the opportunity to respond) and participants were instructed to copy it. Results showed that participants who received the correction following an error answered more questions correctly than participants in the other condition. Similarly, Markle (1990) discusses the ineffectiveness of training methods that prompt students to copy the correct response. She asserts that having students simply copy the correct response does not result in “meaningful” behavior, because they can copy the answer without paying much attention to the response or discrimination.

In addition to the form of feedback, there is also controversy in regard to the timing of feedback. Feedback can occur either before or after the response. Tosti (1978) recommends that corrective feedback be delivered before the next opportunity to

respond, as opposed to following an incorrect response when teaching a new skill. He comes to this conclusion by delineating two different types of feedback: motivational and formative. These types of feedback also have different functions.

Motivational feedback, typically referred to as praise, serves a reinforcing function. It aims to increase the occurrence of an already existing behavior. As such, this type of feedback is delivered after the behavior occurs. Formative feedback, typically referred to as corrective feedback, serves to change the form or stimulus control of behavior. In practice, formative feedback has been known to occur following an incorrect response and before the opportunity to emit the next response. Tosti (1978) recommends that this type of feedback should occur only before the opportunity to respond. This way, the two functions of feedback remain separate. Otherwise, formative feedback (correction) becomes enmeshed with motivational feedback (praise), compromising the effectiveness of each. Simply put, “motivational feedback should be immediately given. Formative feedback should be immediately useful” (Tosti, 1978, p. 20).

The technological boom in education demands that researchers discover the most advantageous methods for presenting instructional stimuli. The purpose of this study was to evaluate the effects of three different training methods in teaching individuals to type a specified three-letter response in the presence of a corresponding symbol. The training methods were designed to investigate the effects of the copy response and the timing of the copy response. The copy before (CB) training format required participants to copy the correct response before the opportunity to respond. The guess (G) format provided correction following an incorrect response by displaying

the correct answer. The guess and copy after (GCA) format provided correction following an incorrect response by having participants copy the correct response.

METHOD

Participants

Four adults, ranging from 20-45 years of age participated in the study. Three of them were female, and one was male. All were students at the University of North Texas studying business, English, or behavior analysis.

Setting and Materials

The study took place in Chilton Hall, Room 393B at the University of North Texas. The room contained a table, a chair, and a computer. A Microsoft Visual Basic ® program (Microsoft Corp., http://www.microsoft.com) program was used to run sessions. This presented all instructions, stimuli, and feedback and collected all experimental data. Participants responded by typing on the computer's keyboard and operating the computer's mouse. The experimenter was not in the room during experimental sessions.

Experimental stimuli consisted of symbols from a computer's abstract clip art. There were three sets of ten symbols. Each of the sets contained an equivalent number of shapes (squares, circles, diamonds, etc), and contained symbols with similar fill patterns. Each symbol had a corresponding response that consisted of a consonant-vowel-consonant (CVC) trigram. The experimenter selected trigrams with a low scaled meaningfulness score (see Noble, 1961). The scaled meaningfulness scores of selected trigrams ranged from .00 to .87 (.00 being the least meaningful). Trigrams were assigned to sets so that the first letter of each trigram occurred at the same frequency across sets. For each set, "X" was the first letter for four of the trigrams, and "Q," "Z," and "Y" was the first letter for two of the trigrams. Similar attention was also given to the

second and third letter of the trigrams. However, an exact balance (as was achieved for the first letter) was impossible given the trigrams available with a scaled meaningfulness score of .87 or less.

The three sets of stimuli were then assigned to one of three training formats. Table 1 outlines the stimulus-response pairs within each format, the scaled meaningfulness scores of the trigrams, and the sum of scaled meaningfulness scores for each training format. During training, the stimuli were presented in a random order.

Measures and Data Collection

Dependent measures were cumulative number of correct, incorrect, and skip responses; cumulative duration of training trials; and frequency of error topographies in the presence of particular symbols. The computer collected all data on these measures.

A correct response was defined as the participant typing the trigram that corresponded with the symbol on the screen, and then clicking “OK” or pressing the enter key during testing trials.

An incorrect response was defined as the participant typing one or more letters that did not correspond with the symbol on the screen, typing letters that correspond with the symbol out of order, or typing more or less than three letters, then clicking “OK” or pressing “enter” during testing trials.

A skip response was defined as the participant typing the numeral zero, then clicking “OK” or pressing the enter key during testing trials.

Duration of training trials was defined as the time elapsed from when the training trial began to when it ended. For the copy before (CB) training format, a trial began when a CB training screen was presented, and ended when participants copied the

correct response into the textbox provided and clicked “OK” or pressed the enter key. For The guess (G) training format, a trial began when a G training screen was presented and ended if participants typed the correct response and clicked “OK” or pressed the enter key. If participants typed an incorrect response, the trial ended after the computer advanced to the next screen (which displayed the correct response) and participants clicked “OK” or pressed the enter key. For the guess and copy after (GCA) training format, a trial began when a GCA training screen was presented and ended if participants typed the correct response and clicked “OK” or pressed the enter key. If participants typed an incorrect response, the trial ended after the computer advanced to the next screen (which displayed the correct response) and participants copied the correct response into the textbox provided and clicked “OK” or pressed the enter key.

Procedures

Introduction

Upon arrival to a session, participants were greeted and accompanied to Room 393B. Prior to the first session, the participants were provided with a brief written explanation of the study and signed the informed consent form (see Appendix). The experimenter then launched the program and gave the instruction to “Read the on-screen instructions carefully and follow those instructions.” The experimenter then left the room for the remainder of the session.

During training, 10 consecutive trials were presented for each format. The 10 training trials for each format were always followed by 10 corresponding testing trials. One sub-session included training and then testing each format once. There were 4 sub-sessions per session for approximately the first half of the study. Thereafter, there

were 6 sub-sessions per session. Sessions occurred within three days of one another. The number of sessions for each participant varied (ranging from 6 to 11 sessions), depending on when they achieved mastery criterion. Mastery criterion was defined as 100% accuracy for all 30 symbols across 6 consecutive testing sub-sessions, or 100% accuracy for 10 symbols from one of the training formats across 6 consecutive testing sub-sessions. Session duration varied, ranging from approximately 25 min to 90 min, depending on how quickly participants responded to experimental stimuli. The study took place over a period of 5 wks.

Pretraining

Before the session began, participants were trained in the three main trial formats that would be used throughout the study. This was done to ensure they understood the instructions for each trial format and knew when to use the three different responses (copy, guess, and skip). A different response was required for each trial format.

A copy response was required for trials that instructed participants to “copy the correct response.” Participants copied by typing the modeled trigram into the text box provided. A guess response was required for trials that instructed participants to “guess” if they did not know the correct answer. Participants guessed by typing any three letters into the text box provided. A skip response was required on trials that instructed participants to “type the numeral 0” if they did not know the correct response. Participants skipped by typing the numeral zero on the keyboard. When participants followed instructions by emitting the specified response (either copy, guess, or skip) and

then clicked “OK,” a screen displaying “Great!” appeared, and the computer advanced to the next pre-raining trial.

There were 3 pretraining trials, one for each of the trial formats. Pretraining trials looked identical to experimental trials, except that symbols and trigrams used in pretraining were not used during the experiment. When the pretraining trials were completed, the computer advanced to a screen that notified participants that they would now begin the experiment. Once participants clicked “OK,” the computer advanced to baseline (for the first session), or to training trials (for all subsequent sessions).

Baseline

A trial began when the computer displayed one of the symbols with the following instructions: “Please type the correct answer. Do not guess. If you don’t know it, type the numeral 0.” After participants responded by typing into the textbox, they clicked the “OK” button or pressed the enter key to advance to the next screen.

If the response was correct, a screen displaying “Great!” appeared, and the next trial was presented. If participants typed the numeral 0 or an incorrect three-letter response, the computer advanced to the next trial. If participants typed more or less than three letters, or characters other than letters, they were reminded that the answer is always three letters, and were asked to try again. The trial was then repeated. See Figure 1 for a sample screenshot of a baseline trial. See Figure 2 for a diagram of the baseline procedures.

There were 30 consecutive baseline trials in one baseline sub-session. Each trial contained one of the 30 symbols. Symbols were presented in a recurring order so that a symbol from copy before (CB) was presented, followed by a symbol from guess (G),

followed by a symbol from guess and copy after (GCA), and so forth. There was one baseline sub-session throughout the entire experiment, which occurred during the first session immediately following pretraining. After subjects responded on the 30th baseline trial, the computer immediately advanced to training.

General Procedures: Training

There were three types of training formats participants encountered throughout the study: CB, G, and GCA. Ten consecutive trials were presented for each format. Each trial contained one of the symbols assigned to that training format. The order in which participants received training formats was counterbalanced (see Table 2). After participants responded by typing into the textbox, they clicked the “OK” button or pressed the enter key to advance to the next screen.

Copy Before (CB)

A trial began when the computer displayed one of the symbols with the following instructions: “The correct answer is (e.g., GEF). Please type the correct answer.” If participants responded correctly by copying the correct answer, a screen displaying “Great!” appeared, and the computer advanced to the next trial. If participants typed anything other than the model, they were given the instruction to try again, and the trial was repeated. See Figure 3 for a sample screenshot of a CB trial. See Figure 4 for a diagram of the procedures for CB trials.

Guess (G)

A trial began when the computer displayed one of the symbols with the following instructions: “Please type the correct answer. If you don’t know it, guess.” If participants typed the correct response, a screen displaying “Great!” appeared, and the computer

advanced to the next trial. If participants typed “0” to skip, they were reminded that they had to guess to continue, and the trial was repeated. If participants typed more or less than three letters, or characters other than letters, they were reminded that they must type three letters to continue, and the trial was repeated. If participants typed any three letters other than the correct answer, the computer advanced to a correction screen. This screen displayed the same symbol and the following text: “The correct answer is (e.g., SAF)”. Once participants clicked “OK” or pressed the enter key, the computer advanced to the next trial. See Figure 5 for a sample screenshot of a G trial. See Figure 6 for a diagram of the procedures for G trials.

Guess and Copy After (GCA)

The screen layout, instructions, and procedures were the same as G trials, but with one additional requirement: a copy response on the correction screen. The correction screen displayed a symbol with the following instructions: “The correct answer is (e.g., HEJ). Please type the correct answer.” If participants responded correctly by copying the correct answer, the computer advanced to the next trial. If participants typed anything other than the correct answer, they were instructed to try again. See Figure 5 for a sample screenshot of a G and GCA trial. See Figure 6 for a diagram of the procedures for G and GCA trials.

Testing

The screen layout, instructions, and procedures were identical to baseline trials. See Figures 1 and 2. There were 10 consecutive testing trials that followed 10 consecutive training trials. The symbols used in testing trials were dependent on which

training format had preceded. For example, testing trials following CB used symbols from CB.

Follow-up

The screen layout, instructions, and procedures were identical to baseline and testing trials (see Figures 1 and 2). There was one follow-up session. It consisted of 30 consecutive trials. Each trial contained one of the 30 symbols. Trials were presented in the same order as baseline. The follow-up session was held 2 wks after participants' final testing session.

Design

The experimental design was an alternating treatment design imbedded in an A-B-A-B design (Barlow & Hernsen, 1984). The three training formats were interspersed with testing trials in order to evaluate the effects of each training format on correct, incorrect, and skip responses, duration of training trials, and response topographies to specific symbols.

RESULTS

Figure 7 displays the number of sub-sessions until criterion (three consecutive sub-sessions at 100%) and cumulative training time for each participant for all training formats. The top panel displays these data for Participant 1, the second panel displays these data for Participant 2, the third panel for Participant 3, and the bottom panel for Participant 4.

For Participant 1, copy before (CB) required 33 sub-sessions to reach criterion. Guess (G) required 36, and guess and copy after (GCA) 33. CB resulted in a total of 44 min, G resulted in 63 min, and GCA 68 min.

For Participant 2, CB required 49 sub-sessions to reach criterion. G required 45, and GCA 42. CB resulted in a total of 48 min, G resulted in 105 min, and GCA 108 min.

For Participant 3, CB required 23 sub-sessions to reach criterion. G required 16, and GCA 14. CB resulted in a total of 21 min, G resulted in 42 min, and GCA 36 min.

For Participant 4, CB did not reach criterion before Sub-session 42. It averaged 93% accuracy on the final three sub-sessions (Sub-sessions 40- 42). G required 42 sub-sessions to reach criterion, and GCA required 34. CB resulted in a total of 28 min, and G and GCA each resulted in 45 min.

Figure 8 shows the performance of Participant 1. The top graph displays the number of correct responses, the second graph displays the number of incorrect responses, and the third graph displays the number of skip responses for each training format across sub-sessions. The bottom graph displays the cumulative amount of time spent in training for each format.

The number of correct responses (top graph) for the three training formats increased rapidly for the first three sub-sessions. CB continued to increase and remained at or above seven correct responses following the 5th sub-session. G and GCA remained relatively level (between three and six correct responses) through Sub-session 12. Around Sub-session 13, G and GCA rapidly increased to 7 correct responses and remained high. CB and GCA reached a steady state of 10 correct responses at Sub-session 31. G continued to fluctuate between 7 to 10 correct responses throughout the duration of the experiment.

The number of incorrect responses (second graph) for the three formats increased in a similar manner for the first three sub-sessions. On Sub-sessions 4 through 8, CB decreased rapidly. G and GCA also decreased, yet performance was more variable. Following Sub-session 11, three or less incorrect responses occurred per sub-session for all formats. CB and GCA reached a steady state of 0 incorrect responses on Sub-session 31, while G continued to fluctuate between 0 to 3 incorrect responses throughout the duration of the experiment.

The number of skip responses (third graph) for the three formats decreased rapidly for the first three sub-sessions. After Sub-session 3, no skip responses occurred in CB. G and GCA gradually decreased until they reached 0. GCA reached 0 at Sub-session 19, and G reached 0 at Sub-session 28.

The cumulative training time (bottom graph) for each format was similar for the first four sub-sessions. At Sub-session 5, CB decreased and continued that trend throughout the experiment. G and GCA also decreased, but more gradually. CB resulted in a total of 44 min. G resulted in 63 min, and GCA resulted in 68 min.

Figure 9 shows all the responses made in testing by Participant 1. The top panel displays responses made to stimuli in the CB training format. The middle panel displays responses made to stimuli in the G format. The bottom panel displays responses made to stimuli in the GCA format. Black boxes denote correct responses, dark grey boxes denote skip responses, and light grey boxes denote incorrect responses. The light grey boxes also contain the error that was emitted.

In the top panel (CB), there were a total of 44 incorrect responses, and 19 different error topographies. Fifteen of the topographies occurred only once or twice. The most frequent error was “ZEH”, which occurred 10 times for stimulus ZEQ. This error was also emitted at follow-up.

In the middle panel (G), there were a total of 61 incorrect responses, and 27 different error topographies. Twenty-one of the topographies occurred only once or twice. The three most frequent errors were “YUJ,” “ZIF,” and “ZUJ.” Each occurred 6 times. The error “QUH” was emitted 4 times for stimulus QIH, and was also emitted at follow-up. For response ZIH, there were 15 errors (with six different topographies). This item was skipped at follow-up.

In the bottom panel (GCA), there were a total of 53 incorrect responses, and 21 different error topographies. Thirteen of the topographies occurred only once or twice. “XIY” was the most frequent error, and it also occurred at follow-up. For response XIH, there were 18 errors (with eight different topographies). One of these errors also occurred at follow-up.

Figure 10 shows the performance of Participant 2. The number of correct responses (top graph) for all formats began around eight or nine. At Sub-session 15,

correct responding for all formats decreased to around five or six, and remained at this level through Sub-session 25. At Sub-session 26, all formats increased to around seven or eight correct responses. After Sub-session 35, the number of correct responses for all formats remained at or above seven.

The number of incorrect responses (second graph) for all formats ranged from 0 to 4 through Sub-session 23. Around Sub-session 24, an increasing trend became apparent, especially for G. Higher levels for all formats were observed through Sub-session 35. Thereafter, all formats decreased, and remained at or below three incorrect responses for the duration of the experiment.

The number of skip responses (third graph) remained at or below one for all formats for the first fifteen sub-sessions. Higher rates were observed on Sub-sessions 16 through 24. This was especially true for CB, which ranged from 0 to 5 skip responses, while G and GCA ranged from 0 to 3. Following Sub-session 25, only 4 skip responses occurred for the remainder of the experiment.

The cumulative training time (bottom graph) for each format was similar for the first four sub-sessions. At Sub-session 5, CB decreased and continued that trend throughout the experiment. G and GCA also decreased, but more gradually. CB resulted in a total of 48 min. G resulted in 105 min, and GCA resulted in 108 min.

Figure 11 shows all the responses made in testing by Participant 2. In the top panel (CB), there were a total of 62 incorrect responses, and 19 different error topographies. Eleven of the topographies occurred once or twice. The most frequent error was “ZEJ,” which occurred 10 times for the stimulus YUJ and once for ZOJ. There were no errors made at follow-up.

In the middle panel (G), there were 76 total incorrect responses, and 25 different error topographies. Fifteen of the topographies occurred once or twice. The most frequent error was “XIH,” which occurred 14 times across 4 different stimuli. For response XIY, there were 25 errors (with four different topographies). This item was skipped at follow-up.

In the bottom panel (GCA), there were a total of 51 incorrect responses, and 28 different error topographies. Twenty-one of the topographies occurred once or twice, and none were emitted more than five times. “XUH” and “QUJ” were the most frequent errors, each occurring 5 times. These topographies were the only two errors emitted at follow-up.

Figure 12 shows the performance of Participant 3. The number of correct responses (top graph) increased rapidly during the first five sub-sessions for CB and GCA, while G increased more gradually. Following Sub-session 5, CB and GCA remained at or above eight correct responses. G gradually increased through Sub-session 10, when it reached 9 correct responses and remained high. CB reached a steady state of 10 correct responses at Sub-session 21. G and GCA continued to fluctuate around nine correct responses for the duration of the experiment.

The number of incorrect responses (second graph) for GCA decreased rapidly within the first three sub-sessions. CB and G gradually decreased within the first 10 sub-sessions. For all formats, one or less incorrect responses were emitted per sub-session following Sub-session 10. For CB, a steady state of 0 incorrect responses was established on Sub-session 21. G and GCA continued to fluctuate between 0 and 1 incorrect responses throughout the duration of the experiment.

The number of skip responses (third graph) for CB and G decreased to 1 within the first four sub-sessions. GCA jumped from 2 to 6 on the second sub-session, and then rapidly decreased. CB and GCA reached a steady state of 0 skip responses on Sub-session 8, and G reached a steady state on Sub-session 12.

The cumulative training time (bottom graph) for each format was similar for the first two sub-sessions. At Sub-session 3, CB decreased and continued to decrease throughout the experiment. G and GCA began to decrease around Sub-session 10. CB resulted in a total of 21 min. G resulted in 42 min, and GCA resulted in 36 min.

Figure 13 shows all the responses made in testing by Participant 3. In the top panel (CB), there were a total of 21 incorrect responses, and 11 error topographies. Nine of the topographies occurred once or twice. The most frequent error was “XIF”, which occurred 6 times for response XIJ. For response ZEQ, 4 errors were emitted, and this item was skipped at follow-up.

In the middle panel (G), there were a total of 38 incorrect responses, and 22 different error topographies. Eighteen of the topographies occurred once or twice, and none occurred more than four times. For response QIH, there were 8 errors emitted, each with a different topography. One of the topographies also occurred at follow-up.

In the bottom panel (GCA), there were a total of 19 incorrect responses, and 15 different error topographies. All of the topographies occurred only once or twice. No errors occurred at follow-up.

Figure 14 shows the performance of Participant 4. The number of correct responses (top graph) for GCA increased from 0 to 6 within the first several sub-sessions, and CB and G increased from 0 to 3. GCA continued to steadily increase,

while CB and G increased more gradually. GCA and reached a steady-state of 10 correct responses on Sub-session 32. CB and G reached 10 correct responses around Sub-session 40, although steady responding at this level was not maintained.

The number of incorrect responses (second graph) began at 5 or 6 for all training formats. All formats gradually decreased in a similar manner throughout the experiment, although there was a fair amount of variability. GCA reached a steady state of 0 incorrect responses on Sub-session 32. CB and G did not maintain a level of 0 incorrect responses.

The number of skip responses (third graph) was variable throughout the experiment for all training formats. Generally, CB had a higher number of skip responses, followed by G, and lastly GCA. GCA reached a steady state of 0 skip responses at Sub-session 31. G reached a steady state at Sub-session 37, and CB reached a steady state at Sub-session 38.

The cumulative training time (bottom graph) for G and GCA was nearly the same for the first four sub-sessions. CB took less time than the other formats. G and GCA gradually decreased in a similar manner throughout the experiment, while CB remained relatively the same. G and GCA each resulted in a total of 45 min, and CB resulted in 28 min.

Figure 15 shows all the responses made in testing by Participant 4. In the top panel (CB), there were a total of 110 incorrect responses, and 51 different error topographies. Thirty-nine of the topographies occurred once or twice. The most frequent error was “XEJ,” which occurred 8 times for response XEF, and 3 times for two other

stimuli. All of the errors emitted at follow-up were topographies that occurred during sub-sessions.

In the middle panel (G), there were a total of 97 incorrect responses, and 54 different error topographies. Forty-two of the topographies occurred only once or twice. The most frequent error was “YEJ,” which occurred 6 times for response YOJ, and once for another stimulus. All of the errors emitted at follow-up were topographies that occurred during sub-sessions.

In the bottom panel (GCA), there were a total of 72 incorrect responses, and 42 different error topographies. Thirty-three of the topographies occurred only once or twice, and none occurred more than five times. Two of the three errors at follow-up also occurred during sub-sessions. For response ZUJ, no errors occurred following the 14th sub-session. However, an error occurred at follow-up.

DISCUSSION

The results of this study show that the copy before (CB) training format is a more efficient training method than guess (G) and guess and copy after (GCA) formats. This is evidenced by examining the number of correct responses in conjunction with the cumulative training time. In addition, CB also produced fewer error topographies than the other formats.

With respect to accuracy, it was unclear which format was most effective. For Participant 1, CB had a higher number of correct responses. For Participant 2, there was little difference between the three formats. For Participant 3, CB and GCA had a similar number of correct responses, which were higher than G. For Participant 4, GCA produced a higher number of correct responses. Given this information, one might conclude that the formats were similarly effective. This conclusion seems valid, because number or percent correct has become a universal measure to assess performance. However, this measure only represents one dimension of responding. Other measures should also be examined to convey a more complete representation of performance.

One such measure is the time students spend during training to achieve a high degree of accuracy. A unique dependent variable of this study was cumulative training time. Overall, although the accuracy of responding was very similar across participants and formats, the cumulative training time was clearly different. For CB, the cumulative training time was significantly lower than the other formats. In fact, CB was 1.6 times faster than the next quickest training format for Participant 1. It was 1.4 times faster for Participant 2, 1.7 times faster for Participant 3, and 2.2 times faster for Participant 4. This worked out to be as much as one hour faster for one participant. The superiority of

CB would not have been revealed without examining additional dimensions of participant responding.

The CB training format also produced fewer error topographies for most participants. For Participant 1, CB produced 19 different error topographies, whereas G produced 27, and GCA produced 21. For Participant 2, CB produced 19 different error topographies, G produced 25, and GCA 28. For Participant 3, CB produced 11 different error topographies, G produced 22, and GCA 15. For Participant 4, CB produced 51 different error topographies, G produced 54, and GCA 42. These outcomes can be accounted for by analyzing the training procedures. The G and GCA formats required participants to guess the correct response during training. Because participants had to guess before they were shown the correct response, they emitted many more errors. This was especially true early in the experiment, before participants had been exposed to the correct response on repeated occasions. As a result of guessing, a large number of errors were emitted in the presence of the experimental stimuli, creating a large repertoire of error topographies. On the other hand, CB did not require participants to guess. Rather, they were shown the correct response and prompted to copy it. Therefore, the only response emitted during CB training was the correct response. This limited response variability and subsequently, a small repertoire of error topographies was created.

It is important to note that although participants did not guess during CB training, they still guessed during testing. All participants guessed (some more than others) rather than emit the skip response during testing when they did not know the correct answer. This occurred during testing for all formats. Just as in G and GCA training,

guessing during testing resulted in emitting errors in the presence of experimental stimuli, which served to create a repertoire of error topographies. Hence, for CB, the repertoire of error topographies was established as a result of guessing during testing. For G and GCA, the already existing repertoire (established during training) was further expanded when participants guessed during testing.

The copy response itself, regardless as to when it occurred during training also appeared to have an effect on error topographies. The CB and GCA formats both included a copy response, whereas G did not. G produced the largest number of error topographies for Participants 1, 3, and 4. For Participant 2, GCA produced the largest number of topographies. These data indicate that the copy response following an incorrect response (GCA) also helped to narrow the error repertoire, although not as much as requiring the copy response before (CB).

The results of this experiment also demonstrate the detrimental effects of repeated error practice. For all formats, some of the error topographies established by guessing during sub-sessions reappeared later in the experiment. The reoccurrence of errors was orderly as opposed to random. When errors were emitted on the 2-wk follow-up, most of them had previously occurred during sub-sessions. For Participant 1, 5 of the 6 errors at follow-up were a topography that also occurred during sub-sessions. The other error (“XIC”) was for the response XOC. Because “I” and “O” are next to each other on the keyboard, this error was considered a mistype. For Participant 2, all 3 errors at follow-up occurred in sub-sessions. For Participant 3, 1 of 2 errors at follow-up occurred in sub-sessions. The other error (“QUJ”) was a topography that occurred during sub-sessions for another format. For Participant 4, 12 of 13 errors at follow-up

occurred in sub-sessions. The other error (“XOY”) was a topography that occurred during sub-sessions for another format. Errors at follow-up occurred even after participants had been responding correctly to the stimulus for quite some time during sub-sessions. To illustrate, for Participant 2, the error “XIH” occurred at follow-up for stimulus ZIH, even though it was 100% correct throughout sub-sessions. “XIH” was an error that occurred frequently for other stimuli during sub-sessions.

The persistence of specific error topographies during follow-up suggests that those topographies were under good stimulus control even though they were never reinforced. Control likely developed as a result of emitting errors in the presence of experimental stimuli. As discussed previously, errors were emitted in two different contexts: during G and GCA training when participants were prompted to guess, and for all formats during testing when participants guessed rather than skipped when they did not know the correct response. Because these opportunities to guess typically resulted in an error, and the error was emitted in the presence of experimental stimuli (which likely established stimulus control), procedures should be designed that limit the opportunity to guess. This experiment attempted such an arrangement by designing the CB format and offering a skip response during testing. CB was successful in limiting guessing during training. However, offering a skip response during testing was not sufficient, as all participants emitted errors during testing for all formats. Further research should investigate methods to properly train the skip response in order to reduce guessing during testing.

The timing of the copy response in the CB training format follows closely the recommendation of Tosti (1978). He suggested that feedback should be delivered

before the next opportunity to respond, as opposed to following an incorrect response when teaching a new skill. As previously mentioned, CB was more efficient than other formats because it produced accurate responding in a fraction of the time and constrained error topographies. To the extent that error topographies are constrained, the procedures approximate errorless learning. Errorless procedures have also been documented in the literature as an effective teaching method (see Sidman, 1985). The results of this study support previous findings by Roberts (1997), who found that providing feedback before the opportunity to respond was more effective than providing feedback following an incorrect response when teaching a novel task.

In addition to the timing of feedback, the other issue at hand is what students do while receiving feedback. The CB and GCA formats both included a copy response during feedback, whereas G simply displayed the correct response. Previous researchers have found that requiring individuals to practice the correct response following an error was more effective than presenting corrective feedback without the opportunity to practice (Barbetta et al.,1993; Drevno et al.,1994). The present study replicated these findings. Both CB and GCA were more effective than G. However, the present study also suggests that although active responding (copying) is important, the timing of that responding is also very important. Although CB and GCA both required a copy response, the results show that CB is superior.

Interestingly, programmed instruction researchers have not favored the copy response. Markle (1990) disputes the usefulness of such responses by asserting that they do not require the individual to emit “meaningful” behavior. This argument is vague, and does little to enhance understanding of the copy response. What she likely means

is that students can emit a copy response without taking time to study the response and discrimination. This could indeed be true. Imagine a student who is given the task to copy from a model ten Spanish words and write them next to their corresponding English word. The student will likely copy the Spanish words next to the English as quickly as possible, then promptly return to what she was doing prior. Now suppose the student is given the same task, but was also informed she would be asked to supply the Spanish word without a model after she completed the copying task. The behavior of the student would then look much different. The student would likely take her time and study the responses much more than she did previously. Her copying would appear slower and deliberate, as opposed to rushed and "meaningless."

In the present experiment, the testing contingency played an important role in establishing the "meaningfulness" of the copy response. When participants did not know of the testing contingency (i.e., on the first training sub-session), their performance was poor. After they experienced testing, their performance improved. Thus, if instructional designers use a copy response, they should make students aware that they will need to supply the response without a model in the future.

Unfortunately, much of educational practice does not coincide with the conclusions drawn from the present study. Most educators rely on a single metric to assess student performance. This is typically number or percent correct. While this measure may appear adequate, the results of this study demonstrate that number correct alone fails to provide a complete representation of performance. Other measures of behavior must be used. Such insufficient measurement can lead to

erroneous conclusions on intervention efficacy, which in turn adversely affects programmatic decision-making.

In addition, teaching procedures similar to the CB format are rare in schools. In fact, it is quite the contrary. Educational systems appear to value procedures similar to G. Consider the following example. A teacher asks a student to spell the word “cat.” The student responds by saying, “I don’t know.” The teacher then replies, “Well then, guess.” Given this instruction, the student guesses how to spell “cat” and likely emits an error. Considering the adverse effects of guessing documented in the present experiment, it is no surprise that the student would continue to have difficulty spelling.

Computer-based instructional programs that are designed similar to CB could be used to teach a wide range of basic academic skills. This could include, but is not limited to, spelling, sight word reading, and simple math facts. Such programs would greatly enhance a student’s educational experience by limiting the amount of errors emitted, and increasing the speed at which new skills are acquired.

Table 1

Stimulus Sets and Trigram Scaled Meaningfulness Scores (see Noble, 1961) for Copy Before, Guess, and Guess and Copy After Training Formats

	Symbol	Trigram	Scaled Meaningfulness Score
Copy Before		XEF	.60
		XUH	.47
		XOC	.70
		XIJ	.40
		QUJ	.40
		QOJ	.62
		ZEQ	.76
		ZOJ	.17
		YUJ	.66
		YIV	.84
			Total Scaled Meaningfulness: 5.62
Guess		XUJ	.40
		XOZ	.58
		XIY	.43
		XOJ	.00
		QIH	.66
		QUH	.73
		ZIH	.70
		ZEJ	.76
		YOJ	.80
		YIQ	.79
			Total Scaled Meaningfulness: 5.85

(table continues)

Table 1 (*continued*).

	Symbol	Trigram	Scaled Meaningfulness Score
Guess and Copy After		XIF	.60
		XUY	.40
		XIH	.40
		XEH	.52
		QIJ	.62
		QUG	.73
		ZUJ	.73
		ZUF	.80
		YEJ	.47
		YOF	.87
			Total Scaled Meaningfulness: 6.14

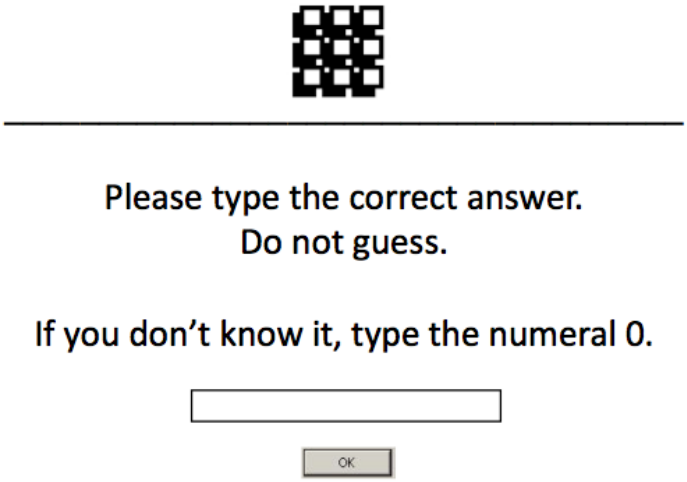

Figure 1. Sample screenshot of baseline, test, and follow-up trials.

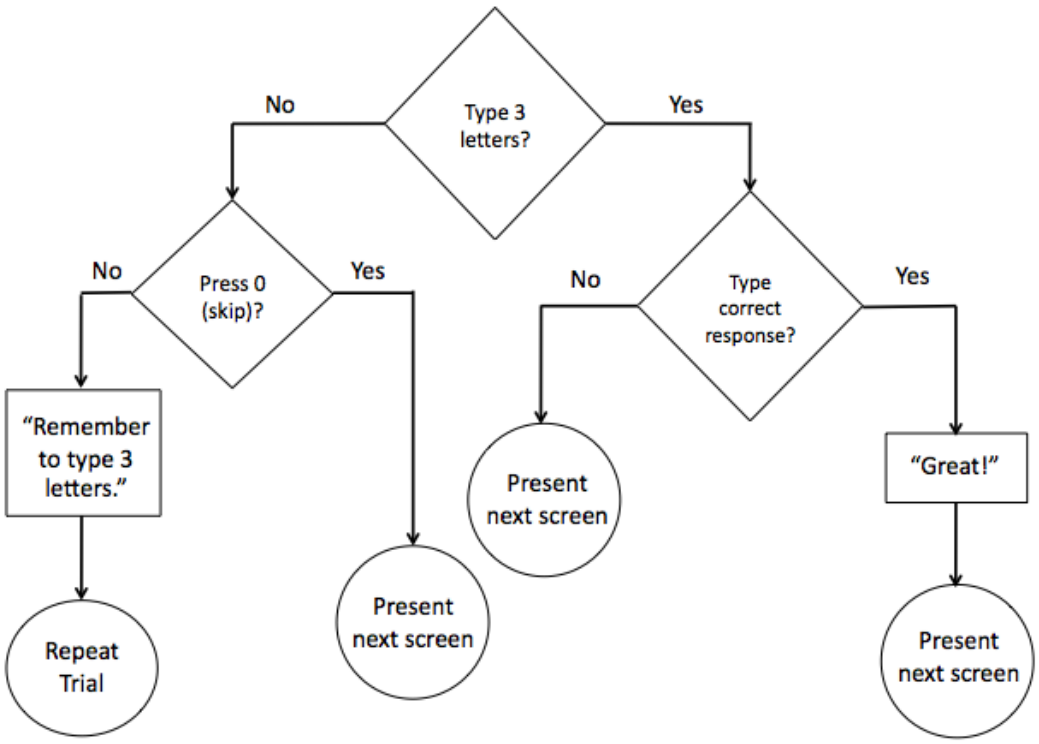

Figure 2. Flowchart of the procedures for baseline, test, and follow-up trials.

Table 2.

Order participants received training formats for each session.

	Subject 1	Subject 2	Subject 3	Subject 4
Session 1	CB, GCA, G	GCA, G, CB	G, CB, GCA	G, GCA, CB
Session 2	GCA, G, CB	G, CB, GCA	CB, GCA, G	GCA, CB, G
Session 3	G, CB, GCA	CB, GCA, G	GCA, G, CB	CB, G, GCA
Session 4	CB, GCA, G	GCA, G, CB	G, CB, GCA	G, GCA, CB
Session 5	GCA, G, CB	G, CB, GCA	CB, GCA, G	GCA, CB, G
Session 6	G, CB, GCA,	CB, GCA, G	GCA, G, CB	CB, G, GCA
Session 7	CB, GCA, G	GCA, G, CB	N/A	G, GCA, CB
Session 8	GCA, G, CB,	G, CB, GCA	N/A	GCA, CB, G
Session 9	G, CB, GCA,	CB, GCA, G	N/A	CB, G, GCA
Session 10	N/A	GCA, G, CB	N/A	N/A
Session 11	N/A	G, CB, GCA	N/A	N/A

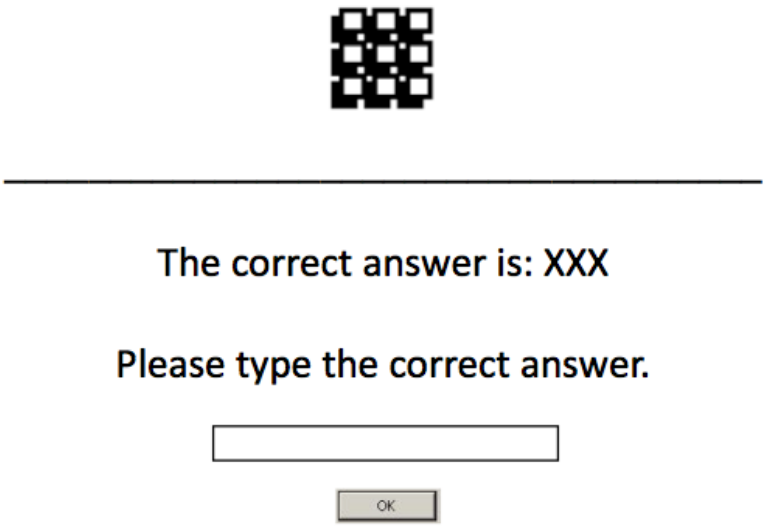

Figure 3. Sample screenshot of a copy before training trial.

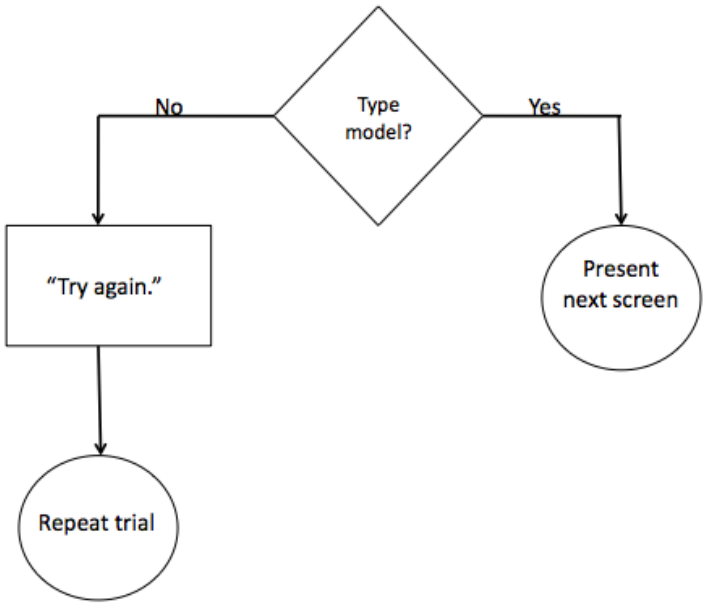

Figure 4. Flowchart of procedures for a copy before training trial.

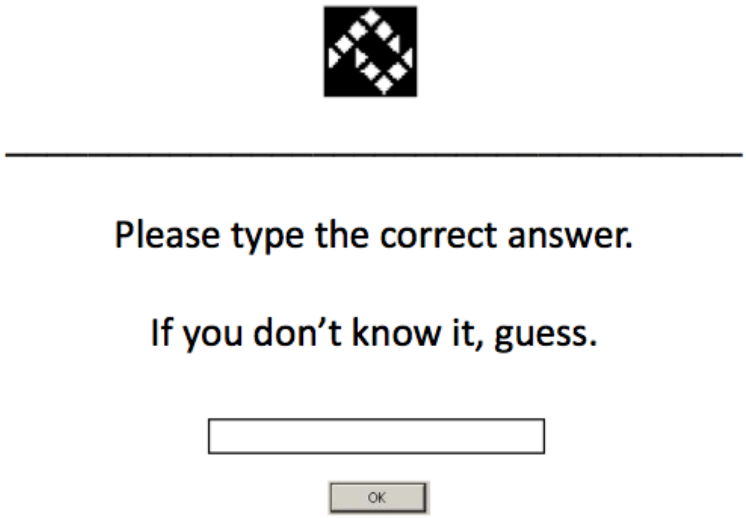

Figure 5. Sample screenshot of a guess and guess and copy after training trial.

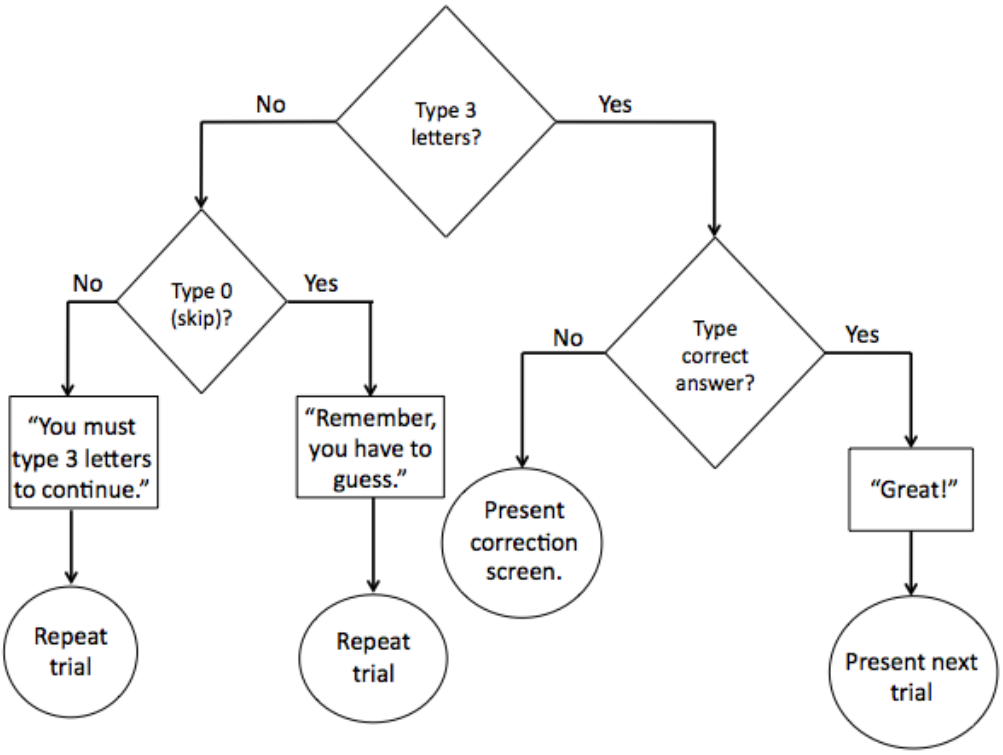

Figure 6. Flowchart of procedures for a guess and guess and copy after training trial.

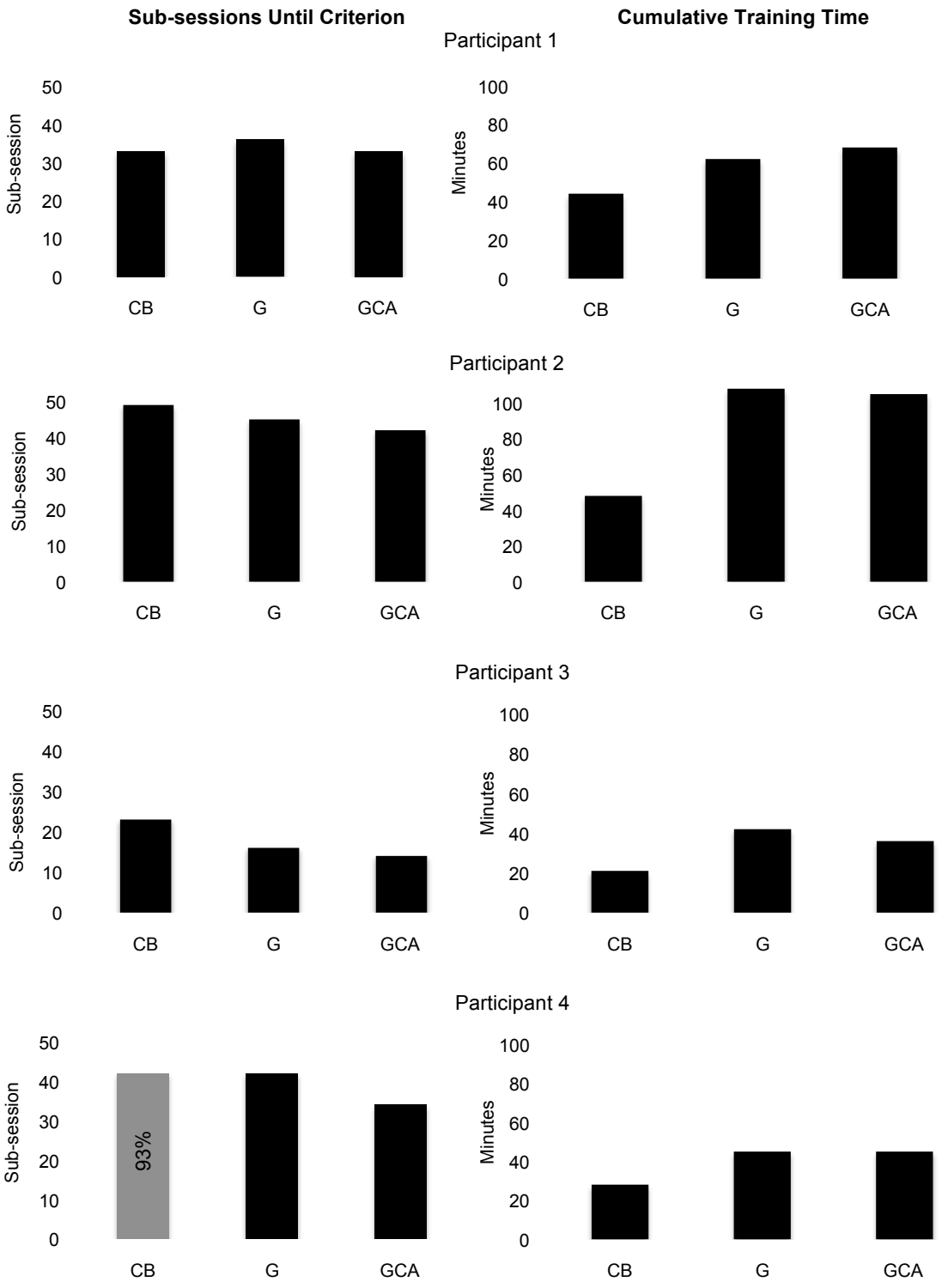

Figure 7. Sub-sessions until criterion and cumulative training time for each participant for all training formats. All but one training format reached criterion (see CB for Participant 4). This format reached an average of 93% accuracy on the final three sub-sessions.

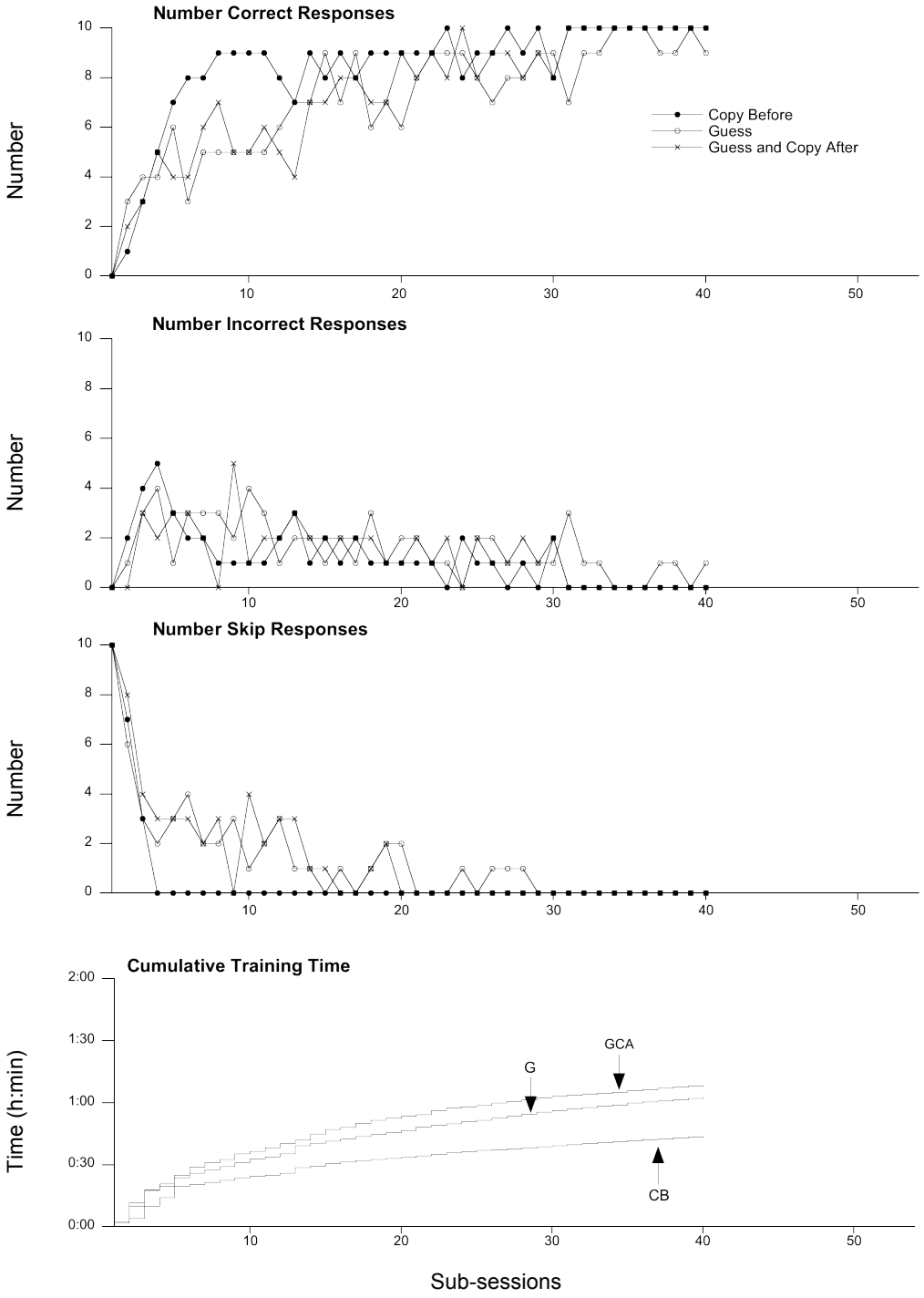

Figure 8. Number of correct, incorrect, and skip responses, and cumulative training time across sub-sessions for Participant 1 during copy before (CB), guess (G), and guess and copy after (GCA) training formats.

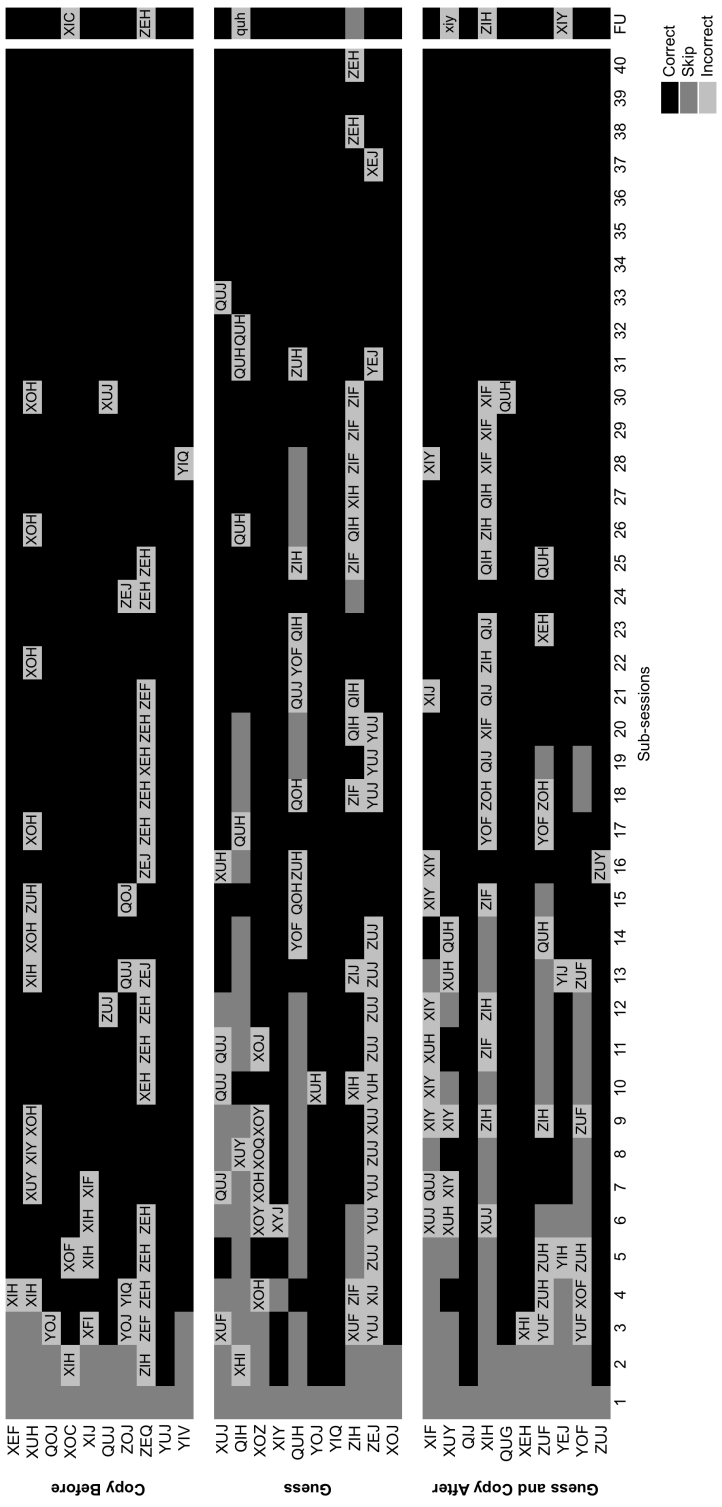

Figure 9. Responses for Participant 1 during baseline (BL), testing sub-sessions, and follow-up (FU) for copy before, guess, and guess and copy After. Black denotes a correct response. Dark grey denotes a skip response. Light grey denotes an incorrect response. Topography of incorrect responses are specified inside the light grey boxes.

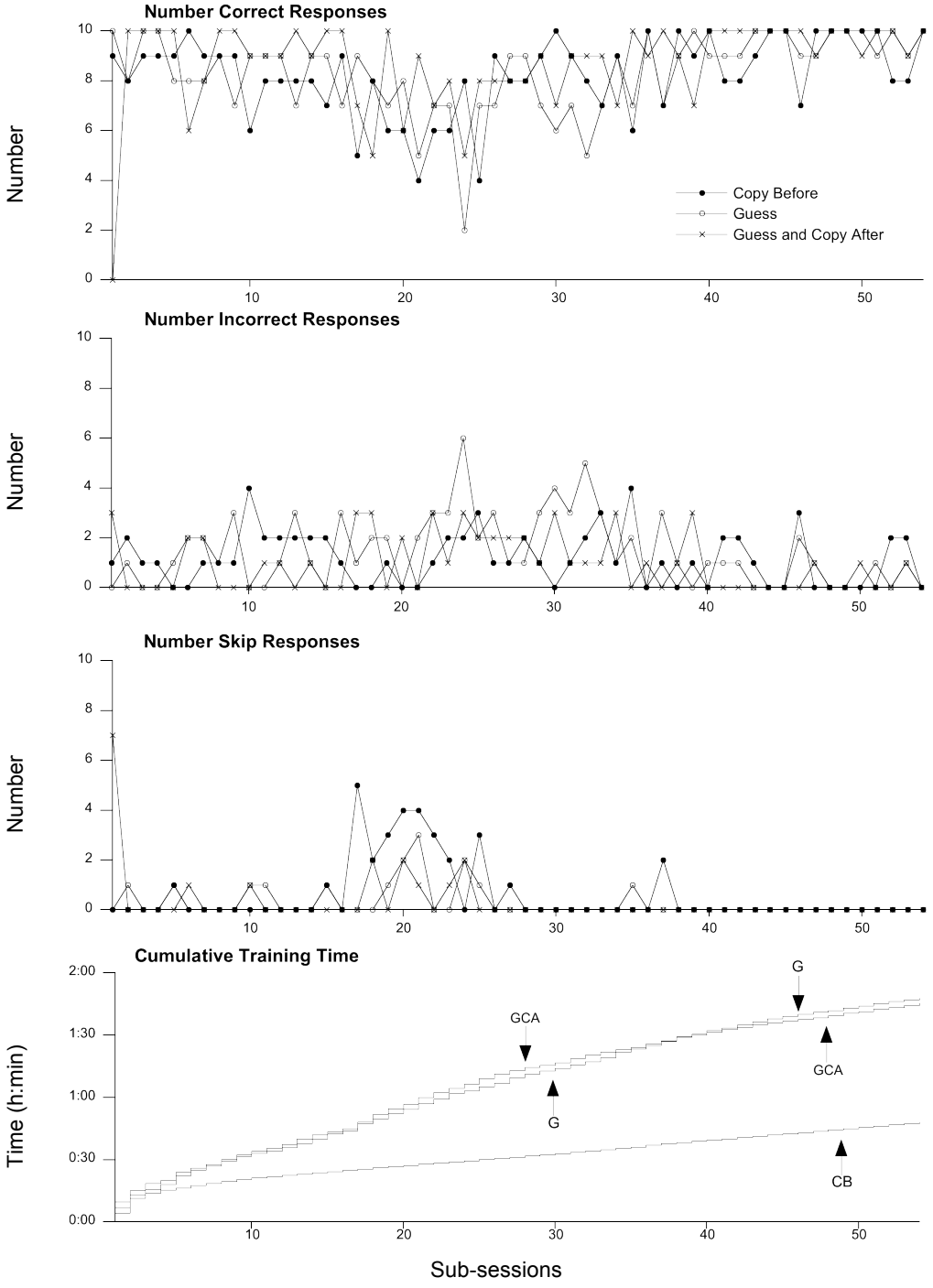

Figure 10. Number of correct, incorrect, and skip responses, and cumulative training time across sub-sessions for Participant 2 during copy before (CB), guess (G), and guess and copy after (GCA) training formats.

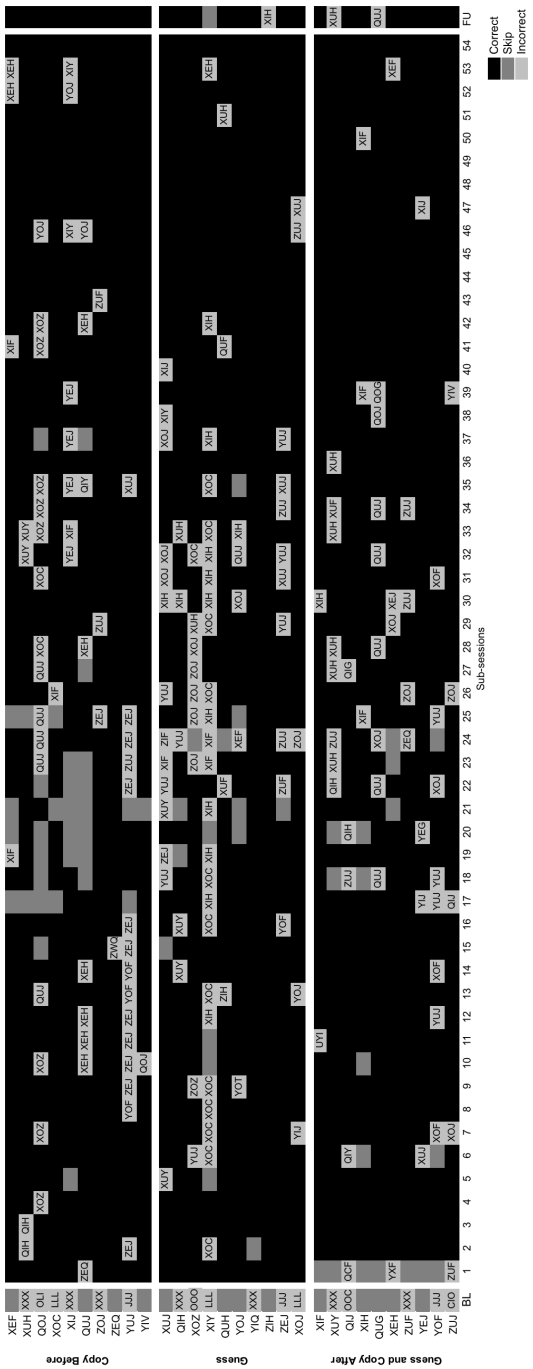

Figure 11. Responses for Participant 2 during baseline (BL), testing sub-sessions, and follow-up (FU) for copy before, guess, and guess and copy after. Black denotes a correct response. Dark grey denotes a skip response. Light grey denotes an incorrect response. Topography of incorrect responses are specified inside the light grey boxes.

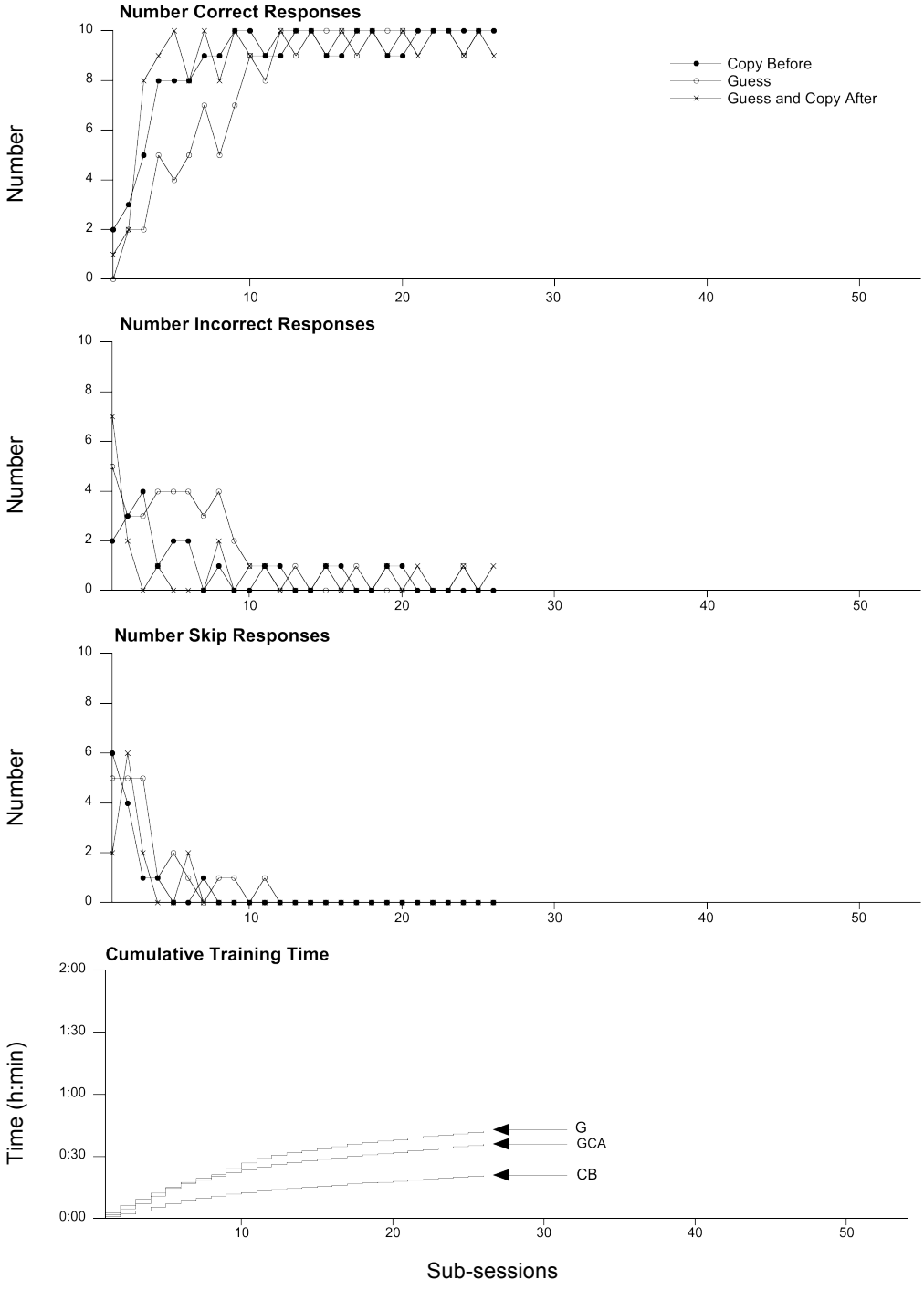

Figure 12. Number of correct, incorrect, and skip responses, and cumulative training time across sub-sessions for Participant 3 during copy before (CB), guess (G), and guess and copy after (GCA) training formats.

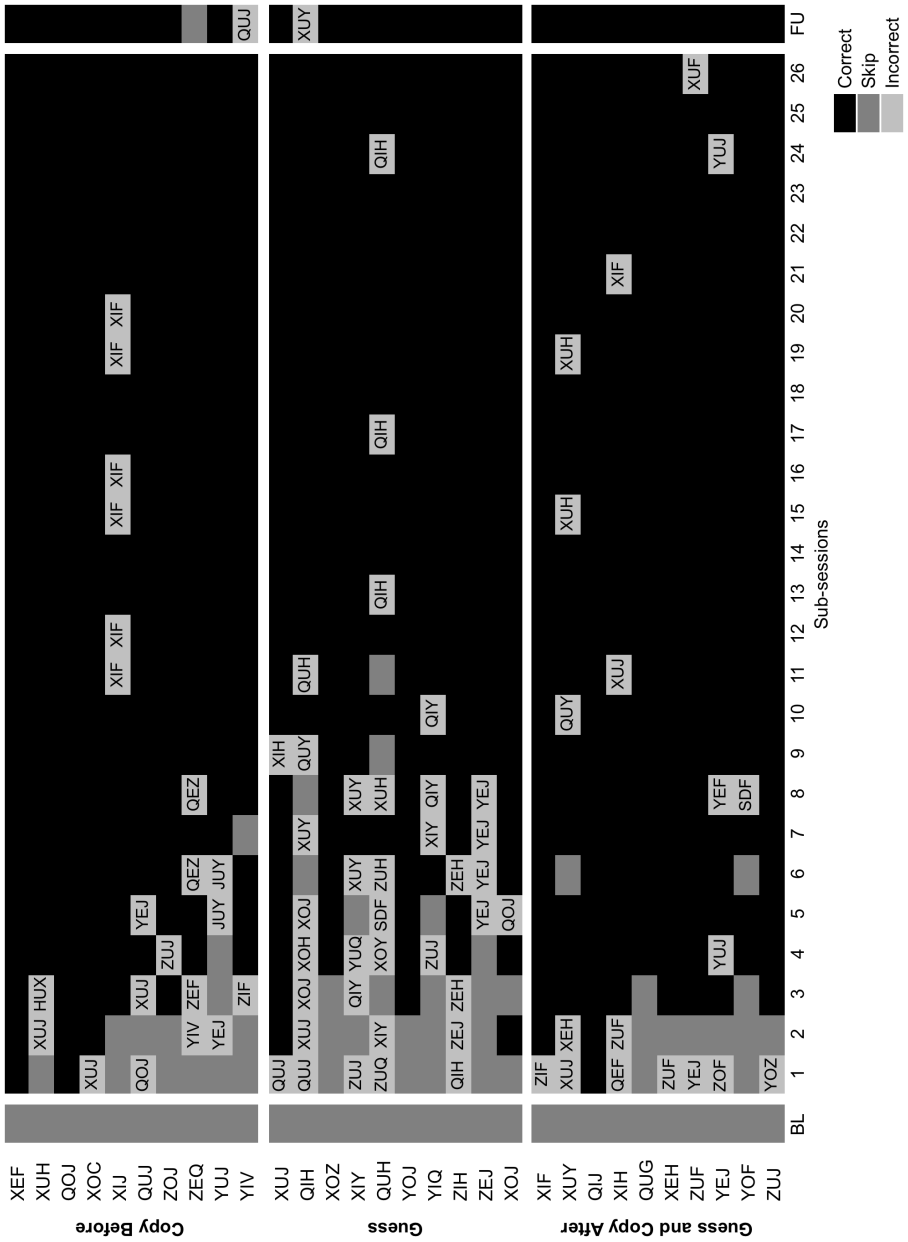

Figure 13. Responses for Participant 3 during baseline (BL), testing sub-sessions, and follow-up (FU) for copy before, guess, and guess and copy after. Black denotes a correct response. Dark grey denotes a skip response. Light grey denotes an incorrect response. Topography of incorrect responses are specified inside the light grey boxes.

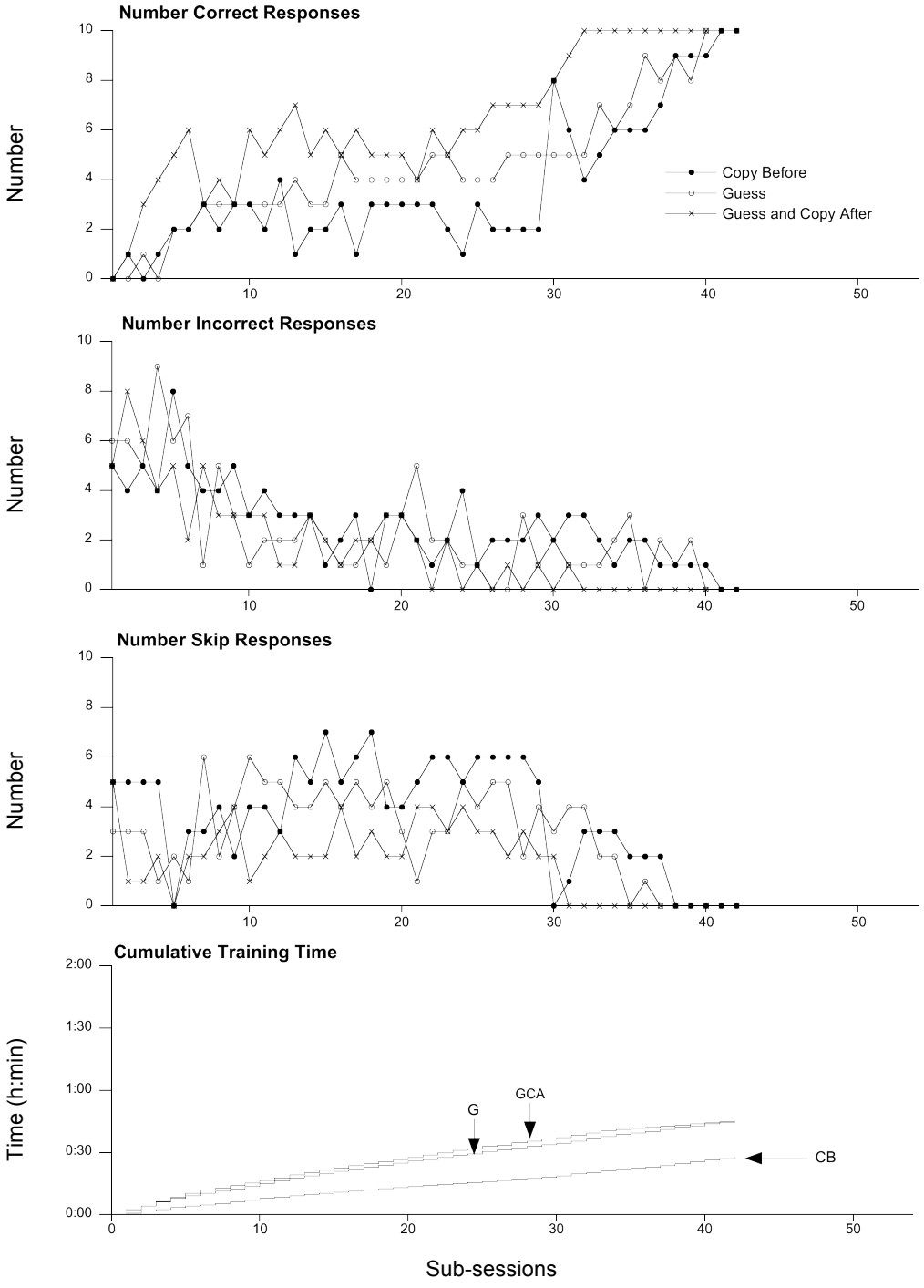

Figure 14. Number of correct, incorrect, and skip responses, and cumulative training time across sub-sessions for Participant 4 during copy before (CB), guess (G), and guess and copy after (GCA) training formats.

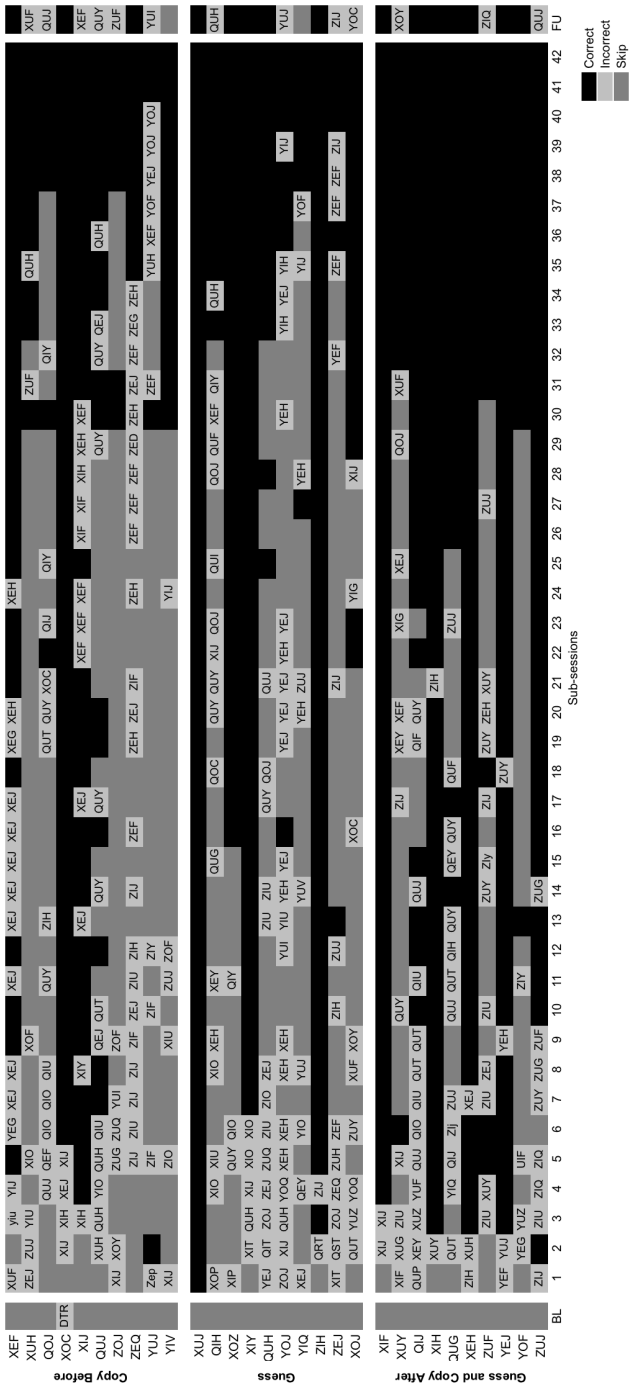

Figure 15. Responses for Participant 4 during baseline (BL), testing sub-sessions, and follow-up (FU) for copy before, guess, and guess and copy after. Black denotes a correct response. Dark grey denotes a skip response. Light grey denotes an incorrect response. Topography of incorrect responses are specified inside the light grey boxes.

APPENDIX

INFORMED CONSENT

University of North Texas Institutional Review Board
Informed Consent Form

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: The effects of errorless and errorful instruction on acquisition rate and retention.

Principal Investigator: Jesús Rosales-Ruiz, University of North Texas (UNT), Department of Behavior Analysis.

Purpose of the Study: You are being asked to participate in a research study to evaluate and compare the effects of three different training methods in teaching subjects to type a specified 3-letter response in the presence of a corresponding symbol.

Study Procedures: The computer will ask you to type 3 letters when a symbol is presented on the screen. You can either guess or skip. Each session will last approximately 45 minutes, and there will be approximately 5 sessions. The entire study will take about 4 hours of your time.

Foreseeable Risks: No foreseeable risks are involved in this study.

Benefits to the Subjects or Others: This study is not expected to be of any direct benefit to you, but the results of this study will greatly contribute to the study of human learning.

Compensation for Participants: You will receive $5.00 for each session attended. Payment will be distributed in full at the completion of the study.

Procedures for Maintaining Confidentiality of Research Records: All records (including signed consent forms) will be kept in a locked filing cabinet. No documents will be posted on the internet. You will be given a pseudonym that will be used when referring to your data, and that pseudonym will be maintained throughout the course of research. All personally identifiable data will be marked with your pseudonym (as opposed to your name) and will remain in a locked filing cabinet. Personally identifiable data will not be disclosed to anyone other than Jesus Rosales-Ruiz and Sarah Pinkelman. The confidentiality of your individual information will be maintained in any publications or presentations regarding this study, including the master's thesis defense.

Questions about the Study: If you have any questions about the study, you may contact Jesus Rosales-Ruiz at telephone number 940-565-2559.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights:

Your signature below indicates that you have read or have had read to you all of the above and that you confirm all of the following:

- Sarah Pinkelman has explained the study to you and answered all of your questions. You have been told the possible benefits and the potential risks and/or discomforts of the study.

- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You have been told you will receive a copy of this form.

_____
Printed Name of Participant

_____ _____
Signature of Participant Date

For the Principal Investigator or Designee:
I certify that I have reviewed the contents of this form with the subject signing above. I have explained the possible benefits and the potential risks and/or discomforts of the study. It is my opinion that the participant understood the explanation.

_____ _____
Signature of Principal Investigator or Designee Date

REFERENCES

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology 62(2),* 148-156.

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist, 40*(4), 193-197.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26,* 111-119.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon.

Drevno, G. E., Kimball, J. W., Possi, M. K., Heward. W. L., Gardner III, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and and generalization of science vocabulary by elementary students: A systematic replication. *Journal of Applied Behavior Analysis, 27*, 179-180.

Geller, D. (1996). *Knowledge-of-correct-response vs. copying-of-correct-response: A study of discrimination learning*. (Unpublished master's thesis). University of North Texas, Denton, TX.

Markle, S. M. (1990). *Designs for instructional designers* (3rd ed). Champaign, Illinois: Stipes.

Noble, C. E. (1961). Measurements of association value, rated associations, and scaled meaningfulness for the 2100 CVC combinations of the English alphabet. *Psychological Reports, 8,* 487-521.

Roberts, P. (1997). *Should corrective feedback come before or after responding to establish a "new" behavior?* (Unpublished master's thesis). University of North Texas, Denton, TX.

Sidman, M. (1985). *Errorless learning and its significance for teaching the mentally handicapped.* Paper presented at the annual meeting of the Psychological Society of Ribeirao Preto, Brazil.

Skinner, B. F. (1968). *The technology of teaching.* Englewood Cliffs, New Jersey: Prentice Hall.

Tosti, D. T. (1978). Formative feedback. *NSPI Journal, 19,* 19-21.