IRis

Northeastern University

Applied Behavioral Analysis Master's Theses

Bouvé College of Health Sciences

January 01, 2010

The effects of varying levels of procedural integrity during prompting on conditional discrimination performance

Meghan Mitchell
Northeastern University

Recommended Citation

Mitchell, Meghan, "The effects of varying levels of procedural integrity during prompting on conditional discrimination performance" (2010). *Applied Behavioral Analysis Master's Theses.* Paper 16. http://hdl.handle.net/2047/d20000123

This work is available open access, hosted by Northeastern University.

The Effects of Varying Levels of Procedural Integrity during Prompting on Conditional Discrimination Performance

A Thesis Presented

by

Meghan Mitchell

The Department of Counseling and Applied Educational Psychology

In partial fulfillment of the requirements

For the degree of

Masters of Science

in the field of

Applied Behavioral Analysis

Northeastern University

Boston, MA

March 2010

NORTHEASTERN UNIVERSITY

Bouve College of Health Sciences Graduate School

Thesis Title: The Effects of Varying Levels of Procedural Integrity during Prompting On Conditional Discrimination Performance

Author: Meghan Mitchell

Department: Counseling and Applied Educational Psychology

Approved for Thesis Requirements of Master of Science Degree

_____ _____

Paula Braga-Kenyon

_____ _____

Meca Andrade

_____ _____

Max Jones

ACKNOWLEDGEMENTS

The author expresses gratitude towards Paula Braga-Kenyon, the chairperson of the author's thesis committee, for her constant support through this process. Special thanks are also offered to Meca Andrade and Max Jones, members of the author's thesis committee, for their time and suggestions and to TJ Horvath, Kristina O'Donnell and Katrina Grandelski for their assistance in collecting inter-observer-agreement data.

Table of Contents

The Effects of Varying Levels of Procedural Integrity During Prompting

On Conditional Discrimination Performance

A. Abstract…… …….....5

B. Introduction……6

1. Lack of Procedural Integrity Data in Published Literature……...…….6

2. Possible Experimental Problems Associated with Lack of Data…....……...…7

3. The Effects of Procedural Integrity on Behavioral Programming.…...…….....8

4. The Effects of Procedural Integrity on Prompting…….……8

5. The Effects of Procedural Integrity on Acquisition of Task Analyses.....10

6. The Effects of Procedural Integrity on Conditional Discrimination….…...…11

C. Method…….....14

1. Apparatus……..…..….....17

2. Experimental Design……18

D. Results……..21

E. Discussion…….……...24

F. References…….....29

G. Tables……...31

H. Figure Captions…..…….……..……...35

I. Figures……..36

Abstract

Research has shown that the reliability with which a behavioral intervention is implemented affects the efficacy of that intervention. The present study investigated whether the lack of procedural integrity on prompt delivery can affect the establishment of discriminative repertoires and the emergence of equivalence relations. Three typically developed adults ages 22 to 26 participated. Procedural integrity, within this experiment, involved varying the time at which a prompt was delivered, specifically through a 2 s additional delay applied to the prescribed prompt. During the acquisition phase, each participant was trained to relate 3 sets of arbitrary stimuli using a computer displaying PowerPoint® slides. Each training block was presented at levels of 100%, 83% or 50% procedural integrity. Integrity values were calculated based upon the number of prompting errors presented in each block of slides. Following the acquisition phase, participants were tested for demonstration of transitivity and equivalence relations. Results for 2 of the 3 participants indicated a higher number of errors during training and post-mastery trials when participants were taught relations at less than 100% integrity. These same 2 participants also demonstrated a greater number of sessions to mastery for relations taught with less than 100% integrity. Results for 1 participant, however, indicated that the least amount of sessions to mastery and errors occurred in the blocks taught at the lowest integrity value. The percentage of independent responding was greatest for all participants during blocks taught at 50% integrity, while for 2 of 3 participants, percentage of prompted responses was highest during the 100% integrity condition.

The Effect of Varying Levels of Procedural Integrity during Prompting on Conditional Discrimination Performance

Procedural integrity, defined as the degree to which a treatment or procedure is implemented as described and intended, is an area overlooked in much of the past behavior analytic literature (Billingsly, White, & Munson, 1980). Procedural integrity plays a major role in the acquisition of skills as well as the effectiveness of behavioral interventions designed to reduce the rate or frequency of problem behavior. Yet, in a majority of published studies, this variable is not taken into consideration, or when it is, levels of integrity are not reported or accurately measured.

Gresham, Gansle, and Noell (1993) addressed the issue of the lack of data on procedural integrity by "evaluating the degree to which studies published in Journal of Applied Behavior Analysis (with children as subjects) between 1980 and 1990 monitored or assessed the implementation of independent variables"(p. 258). In this study, the authors assessed all articles greater than three pages, using subjects under 19 years of age, published within the range mentioned above, for the inclusion of integrity data. The authors found that 158 studies met these requirements. The studies which reported on procedural integrity were classified into three categories: monitored (data were not recorded, but were reported to have been collected by experimenters), assessed (where a percentage of integrity was reported) and neither reported nor assessed. Results from the analysis indicated that only 15.8% of the studies reported a percentage of treatment integrity, 75.3% did not report treatment integrity, and 8.9% reported that procedural integrity was monitored but not reported. The authors further reported that only 34.2% of the articles operationally defined the independent variable, a factor which may have not allowed for the collection of data on the independent

variable to occur. Many questions emerged within the field of behavior analysis following Gresham et al.'s review, specifically the possibility that such a lack of data collection or reporting on this important variable may invalidate many of the conclusions reached in the articles included in the study. Without reliable measurement of the independent variable, it is possible that no real functional relationship exists between the dependent and independent variable, and that outside factors could be responsible for the changes observed in a participant's behavior.

According to Baer, Wolf, and Risley (1968), replication is "at the essence of believability", and as such, is a process only achievable in studies in which all procedures, methods, and variables have been accurately described and measured. Peterson, Homer, and Wonderlich (1982) referenced Baer, et al., by stating that "the primary goal of behavioral analysis has been described as going beyond the simple demonstration of changes in behavior to include the demonstration that changes in the target behavior are functionally related to changes in the environment" (p. 477). Unfortunately, the observation and recording of the independent variable proves to be much more difficult within applied behavioral analysis, where a human must administer the treatment package or other independent variables while another person takes data on its application, versus animal or medical studies where a machine may simultaneously deliver the independent variable and record its delivery. Peterson et al. (1982) also stated the specific difficulties involved in observation of independent variable, including such phenomena as observer reactivity, observer bias, observer drift, and observational complexity. While most published studies require that there be a direct measurement of the dependent variable in the study, few are required to report on the delivery of the variable responsible for change in behavior (Peterson

et al., 1982). Within research studies, it is important that data be collected on both variables due to the nature of experimental design. Collecting data on the dependent variable being measured is commonplace, and due to the subjective nature of many variables it is often necessary to provide observers with an operational definition of such variables. In a similar way, experimenters should also define the implementation of treatments or prompting procedures so that data may be collected on the accuracy of implementation.

Neglecting to collect data on the implementation of the independent variable may greatly impact the validity of the findings of many studies. If the independent variable within a study or clinical programming, such as the type of treatment package delivered or prompting procedure used is not regulated and reliably and accurately measured and recorded, there is no tight control for human variance during its implementation. Thus, due to its potential influence on student/ client behaviors, the study of teacher/clinician implementation variables is necessary. Specifically, two main areas of interest with regard to procedural integrity are teaching new skill and decreasing challenging behaviors.

Treatment integrity with regards to behavioral programming has been reported in few studies. In 1999, Vollmer, Roane, Ringdahl, and Marcus compared the effects of different levels of implementation integrity for two separate components of an intervention, reinforcement of appropriate behavior and extinction of problem behavior. Experimenters alternated varying levels of treatment integrity of a functional communication training (FCT) and extinction treatment package across each participant, and the order of treatments was counterbalanced across participants. Different phases of the study utilized implementation of each part of the treatment package (FCT and extinction) at a different level of integrity, and with values ranging between 0% (treatment was never implemented as described) to 100%

(treatment was always implemented as planned). Specific values used were 100%, 75%, 50%, 25% and 0%. Results indicated that following phases conducted with 100% integrity, aberrant behavior decreased to zero or near zero levels and appropriate behavior increased to high rates. When treatment conditions were implemented with less than perfect integrity (75%, 50%, 25% and 0%) following a condition where a treatment was implemented at 100% integrity, treatment effects persisted, and aberrant behavior, although it increased slightly, remained at a low rate when compared to baseline responding (Vollmer et al., 1999). These findings suggest that it is possible that following 100% accurate implementation of a treatment for a set amount of time, decreases in integrity may not affect the dependent variable (aberrant behavior or skill acquisition). Application of these results to a clinical setting in which more qualified staff may train less qualified staff to implement a procedure suggests that the tendency of clinical programming to lose integrity after repeated implementation without retraining may only yield negligible differences in treatment effectiveness.

Another question raised by the Vollmer et al. (1999) study is the effect of lack of procedural integrity within an extinction procedure as compared to a reinforcement procedure. It is possible that varying the integrity of either of these programs could have different consequences. For example, an extinction procedure conducted at less than 100% integrity may function to intermittently reinforce the behavior on extinction, thereby producing a pattern of responding consistent with a variable schedule of reinforcement. Results of this schedule of reinforcement may be detrimental to a treatment package, causing an increase or persistence of the targeted aberrant behavior. In contrast, a reinforcement program such as the delivery of a reinforcer contingent upon every correct response (FR1)

implemented at lower than 100% integrity could be arranged by increasing the ratio requirements for some skill acquisition trials (such as implementing some skill acquisition programs on an FR3). This particular deviation from the prescribed reinforcement procedure (or decrease in procedural integrity) could be seen as beneficial to a client if the rate of correct responding increases.

Treatment integrity has also been explored in a behavioral context with prompting procedures to decrease aberrant behavior. Wilder, Atwell, and Wine (2006) explored the impact of treatment integrity in increasing child compliance via a three-step prompting sequence. The authors demonstrated that compliance occurred at the highest level (70%-100% and 40%-100% of trials) for skills prompted with 100% integrity. Skills taught with 0% prompting integrity resulted in compliance ranging between 0% and 20% for both participants. Compliance values for the skill taught at 50% integrity fell between 0% and 70% compliance, ranging between 0% and 65% compliance for one participant and 30% and 70% compliance for the other. The authors indicated that the results of this study have real world implications in that when interventions are only implemented correctly for half the time, or are not implemented correctly, the behavior being taught, compliance in this case, sharply decreases in frequency, thereby greatly decreasing the effectiveness of the intervention. This situation mimics a natural environment where a child may be subject to one set of contingencies at school and a program may be implemented at 100% in this setting, but when the child is home or in another setting, the intervention may only be implemented sporadically or not at all. In a case like this, the treatment may fail to be effective or may take a much longer period of time to become effective, due in large part to

the lack of integrity of implementation and consistency across teachers, caregivers and settings.

Treatment integrity with regard to prompting during skill acquisition tasks has also been explored to some degree in the literature. In Perry's 2007 unpublished thesis, the author taught children to complete Lego® constructions using most-to-least prompting procedures. The study investigated whether a reduction in integrity of prompt delivery for placing Lego® pieces in sequence or failure to prompt a specific step in the sequence would affect skill acquisition. The purpose of Perry's study was to compare the rates of acquisition of a behavior chain that involved assembling a Lego® construct, when prompts were provided at 100% integrity as compared to 70% integrity. The decreased procedural integrity condition for Perry's study included three different types of errors: the omission of a prompt, a prompt of the wrong Lego® with correct placement, and an out of sequence placement of the Lego® piece. Errors were counterbalanced across sessions and trials for all participants. The results showed that one of the two participants did learn the construct when taught with only 70% integrity, and only mastered the 100% integrity construct three sessions faster than the construct taught at 100% integrity. The other participant, however, did not learn the construct taught with 70% until a reversal was implemented where that construct was taught with 100% integrity. Cumulative error data revealed that with one participant the construct taught with 70% integrity yielded 21 more errors than the construct taught with 100% integrity, and with the other participant 10 more errors were observed with the construct taught at 70% integrity. Perry's study suggests that the effect of lowered integrity values (specifically, integrity of prompting procedures) as it relates to the duration of time required for acquisition is individual to the student, and differences in performance due to lowered

integrity is not universal, as demonstrated by the difference between the study's two participants. Another possible implication of programming taught with less than 100% integrity is with the maintenance of skills, particularly due to the increased possibility for errors during procedures implemented with techniques (including decreased integrity) which do not allow for errorless learning to occur.

Procedural integrity, the degree to which the independent variable is implemented according to its description, is an integral part of research for various reasons, yet in the past was often overlooked or omitted from published studies. Procedural integrity data is integral to research because it ensures that experimental control was present while data were collected on the dependent variable within a study. Measurement of accurate treatment implementation according to guidelines also is imperative within an applied setting for similar reasons. If an implemented intervention is effective, careful observation of the intervention must occur to ensure that the treatment is implemented as written in order to determine if the intervention as written has influenced the rate of behavior or if, alternatively, modification of the intervention has occurred thus allowing for its success. Conversely, if a treatment package or a prompting procedure is not effective with a particular student, in order to ensure that the behavior of the teacher matches the procedure to be implemented as written, procedural integrity data should be collected.

In addition to task analyses, another teaching procedure commonly used in an applied setting is match-to-sample. This technique of teaching is commonly used to develop skills such as increasing receptive language vocabulary, teaching sight words, and categorization, as well as a wide variety of other academic related tasks. Within this procedure, the two most

commonly utilized prompting strategies include prompt fading (Terrace, 1963a) and delayed prompts (Touchette, 1971).

An example of prompt fading involves the gradual, successive fading of prompts such as in a prompting hierarchy of most to least or least to most restrictiveness within a procedure used to teach skill acquisition. An example of the use of a most to least prompting would be a prescription of the most restrictive prompt in the initial phases of acquisition to minimize errors, followed by a less restrictive prompt (such as a tap or light touch) after mastery at earlier prompting levels has occurred. Delayed prompting involves the delivery of prompts at a particular latency from the discriminative stimulus presentation. Progressive delayed prompting involves the gradual increase of the duration of time between the presentation of the discriminative stimulus and the prompt presentation. Presented at a very small or 0 s delay, prompting during initial acquisition sessions utilizing a delayed prompt procedure is continued until the participant demonstrates (through a pattern of accurate responding) that their responding is controlled by the prompt. Following stability of correct responding, the prompt delay is increased gradually, thereby allowing some amount of independence in responding. In this stage of acquisition, participants are presented with the choice to respond immediately or wait for the delivery of the prompt if stimulus control is not present. From an experimental standpoint, progressive delay prompting allows for observation of the moment of stimulus control transfer, whereas prompt fading procedures, “do not reveal the point at which a subject comes under control of the stimuli that will remain when training is complete” (Touchette, 1971, p. 347).

Yates’ unpublished thesis (2008) investigated the prevalence of prompting errors in a classroom during the first phase of her study. To achieve this, she videotaped six staff

members at a residential facility for children diagnosed with autism spectrum disorders administering a match-to-sample procedure, and then used a stopwatch to determine the time of prompt delivery. Analysis of the actual delivery of the prompt from the taped performances to the prescribed delay for each prompt indicated a lack of integrity for all six participants. Yates extended her study to explore the effects of decreased procedural integrity within a prompting procedure on acquisition of known pictures to written Portuguese words (visual-visual stimuli relations) during Phase 2 of her study. Using three typically developed adults, she exposed each participant to three sets of relations (A_{1-3}-B_{1-3}, C_{1-3}-D_{1-3}, and E_{1-3}-F_{1-3}). Each set of relations was taught with a different integrity value (100%, 45% and 0%) and values were counterbalanced across participants to control for sequence effects. Sessions with 100% integrity were prompted as planned using a graduated delay prompting procedure of 0s, 3s and 5s delays. Sessions administered at less than 100% integrity were prompted incorrectly on ten trials (for 45% integrity) or on all trials (for 0% integrity). Prompting errors consisted of a 2 s less or greater delay than the prescribed prompt. The author found that the greatest number of errors and the lowest number of correct independent responses occurred during the 0% integrity condition, while the greatest number of independent responses and the least number of errors occurred during the 100% integrity condition.

The purpose of the current study is to extend Yates' 2008 thesis by examining the impact of degraded procedural integrity values regarding prompt delivery on the learning of both stimulus-stimulus relations as well as the emergence of untrained equivalence relations. More specifically, the current study explored whether or not varied degrees of integrity on the delivery of delayed prompts affected the rate of skill acquisition or the efficacy of this type of prompting, in the context of a matching-to-sample (MTS) task. The study also

examined the effects of varied degrees of procedural integrity upon the formation of transitivity and equivalence relations. The author asked whether variance around a prescribed delay to a prompt compromised the efficacy of that prompting.

Method

Participants and Setting

Three typically developed adults that work as teachers in a residential facility for children diagnosed with autism spectrum disorders participated in the study. Participants ranged in age from 22-26, and each had no previous experience participating in research at their current place of employment. All three individuals were unaware of the study's purpose and were chosen based on their interest and availability to participate. They were familiar with the use of a computer mouse in selecting items on a screen.

Sessions were conducted in an office equipped with a computer, a table and a chair, in addition to other objects or furniture that were not used during this study.

Response Measurement and Inter-observer Agreement

Dependent variables included the relative frequency of correct responses, prompted responses, independent responses, and errors. A correct response was defined as any instance in which the participant selected the comparison stimulus corresponding to the sample stimulus by clicking with the mouse within 5 s of stimulus presentation. A trial was defined as a PowerPoint® slide consisting of one sample stimulus and three comparison stimuli presented simultaneously (Figure 1). A block of trials was defined as a PowerPoint® presentation consisting of either 18 or 36 trials dependent upon whether the participant was learning a single relation or a mixture of two relations. A relation was defined as a pair of

two different arbitrary stimuli which were presented to a participant on a PowerPoint® slide. There were three different relations presented during each acquisition block of trials. Each stimulus which occurred within the same block of trials (as either a comparison or a sample) was given the same letter (ex. A) but a different number (ex. 1, 2 or 3). The number of each stimulus corresponded to the number of the related stimulus to be taught. For example, the stimulus A_1 was presented with the comparison stimuli B_1, B_2 and B_3, and participants were prompted for correctly selecting B_1 in its presence. Relations were labeled based on the order the relation was learned by each participant. In this way, the first set of relations taught to all participants was $A_{1-3} – B_{1-3}$ and the last set of relations learned was $H_{1-3} – I_{1-3}$. These stimuli can be seen in Figures 2-4. Blocks of single relations consisted of six presentations of each sample stimulus (for a total of 18 trials), with the position of comparison stimuli rotating throughout the trials. Data were collected on the selection of the correct stimulus as well as whether the response was prompted or independent. Blocks of two relations (or mixed block relations) included six presentations of each of the six sample stimuli (for a total of 36 slides) with comparison stimuli again presented and rotated on each slide. Prompted responses were those emitted after the onset of a visual prompt. Independent responses were defined as responses that occurred prior to the onset of the visual prompt. One or more observer recorded the data by hand using a data sheet that contained a picture of each of the slides presented in the block of trials (Figure 5). For each trial, the experimenter recorded a "-" for an incorrect response, a "+" for a correct independent response, a "+p" for a correct prompted response, or "NR" for no response. Incorrect prompted responses did not occur during the experiment. A second observer recorded participant's selection in 31% of sessions across participants. Inter-observer agreement was assessed on a trial-by-trial basis by

comparing observer's data sheets. Agreement was calculated by the number of agreements plus disagreements, divided by the total number of intervals and multiplied by 100. Agreement values ranged between 94% and 100% across participants and averaged 99% across participants.

The primary independent variable was the number of prompting errors within each block of trials. A prompting error was defined as the visual prompt (an arrow) presented at two seconds greater than the prescribed prompt for all trials. For example, at an 83% integrity value, there were three errors within the block of 18 trials at 0 s delay and in each error trial the prompt was presented at two seconds instead of at zero seconds following the presentation of the sample stimulus. Sessions with 100% integrity were designed to have zero errors, sessions with 83% integrity were designed to have three errors and sessions with 50% integrity were designed to have nine errors. Data were collected on the accuracy of these prompt delivery times by reviewing the slides and timing the delay of the prompt used for each slide. These data were used to record independent variable integrity. Data on independent variable integrity were collected in 31% of sessions and ranged between 83% and 100% for all blocks of trials and averaged 96% across all participants. Any errors recorded on prompt timings were the result of experimenter error in programming the PowerPoint® slide animation correctly.

Apparatus

The trials were presented on a computer using Microsoft PowerPoint ®. A trial consisted of the presentation of a slide containing a sample stimulus in the upper center of the slide along with three comparison stimuli at the bottom of the slide (Figure 1). Slides were programmed to automatically advance after five seconds or would advance upon the

participant selecting (with the mouse) a comparison stimuli from the array. Each trial within a training block also contained an imbedded or animated arrow (prompt) which indicated which stimuli was correct. Zero second delay trials contained imbedded arrows which were present when the slide was presented. Trials with a delayed prompt (a prompt was presented at greater than 0s) included an animated arrow which appeared from the bottom of the screen and moved up, pointing to the correct stimulus after a specified interval. Blank inter-trial slides were also presented between each trial and lasted for eight seconds.

Baseline and training blocks for single relations consisted of 18 slides. Mixed block training sessions consisted of 36 slides (18 slides from one relation and 18 from the second relation).

Experimental Design

Each participant was trained on three sets of stimuli: A_{1-3}-B_{1-3}-C_{1-3}, D_{1-3}-E_{1-3}-F_{1-3} and G_{1-3}-H_{1-3}-$I_{1-3,}$ at three different levels of integrity: 100%, 83% and 50%. A set of stimuli consisted of two consecutive pairs of relations (A_{1-3}-B_{1-3} and B_{1-3}-C_{1-3}). Levels of integrity were counterbalanced across participants to prevent sequence effects. Lisa's first set of stimuli (A_{1-3}-B_{1-3}-C_{1-3}) was presented with 83% integrity, the second set (D_{1-2}-E_{1-3}-F_{1-3}) with 100% integrity and the third set (G_{1-3}- H_{1-3}-I_{1-3}) with 50% integrity. Jenn was presented with the initial set (A_{1-3}-B_{1-3}-C_{1-3}) with 50% integrity and the second set (D_{1-2}-E_{1-3}-F_{1-3}) at 83%, and the third set (G_{1-3}- H_{1-3}-I_{1-3}) at 100% integrity. Tim was presented first with a set at 100% integrity (A_{1-3}-B_{1-3}-C_{1-3}), the second set (D_{1-2}-E_{1-3}-F_{1-3}) with 50%, and the third set (G_{1-3}- H_{1-3}-I_{1-3}) with 83% integrity (see Table 1).

Procedure

Prior to baseline, the experimenter administered a pretest to participants to assess their ability to respond to a match-to-sample procedure in a PowerPoint® format. The pretest consisted of nine slides presented with a known picture (i.e. an apple) as a sample and three known words (i.e. dog, cat, apple, etc.) as comparison stimuli. No prompts were presented and no indicators were given for correct or incorrect responses. All participants scored 100% on this pretest and this was used to determine their fitness for this study.

Following the pretest condition, participants were administered a baseline to ensure that they did not demonstrate mastery of the relations to be taught. Baseline sessions were conducted without the presentation of prompts or audible feedback for correct responses. Stimuli used during baseline and training blocks were all arbitrary (Figures 2-4) and all participants did not demonstrate mastery of the relations to be taught prior to initiating training sessions.

Three different levels of integrity were implemented for each participant and during teaching of each relation (100%, 83% and 50%). Each level of integrity was associated with one set of relations for each participant. During the 100% integrity level of teaching, all prompts were presented as prescribed across all relations within the set. During the 83% integrity level, three prompts of the 18 were presented incorrectly. Each incorrect prompted trial response was the same across all participants and all sessions, and it was that the prompt was presented at two seconds greater than the prescribed prompt. For example, at an 83% integrity value, there were three errors within the block of 18 trials at 0s delay and in each error trial, the prompt was presented at two seconds instead of at zero seconds.

Training blocks were conducted in the same manner as baseline, with 18 training slides and each training slide followed by a blank slide. However, during training sessions

prompts were presented and correct response selections were followed by a cash register sound. This sound did not have any trade-in value, and thus was only auditory feedback provided to participants during training blocks. Incorrect responses were followed by no sound and there was no correction procedure. All selection responses resulted in progression to the next slide, regardless of whether they were correct. Initially, no instructions were delivered to participants about the type of block being presented (i.e. baseline, training with prompts, or extinction), but instructions were introduced to each participant after they were presented with a second set of stimuli (D_{1-3}-E_{1-3}-F_{1-3}). Instruction delivery was not included in the initial procedure, but was added to allow the participants to more easily differentiate the type of trials presented (baseline, acquisition, post-mastery, etc.). The instructions delivered prior to training sessions were read to the participants. They were delivered as follows: "The slides you will see will each contain a sample stimulus, located at the top of the screen and three comparison stimuli, located at the bottom of the screen. These slides will contain prompts in the form of arrows. If you respond correctly, you will hear a cash register sound." Criteria to increase to the next prompt level was 16/18 (or 32/36 for mixed block sessions) correct responses (prompted or unprompted). Criteria for administering the post mastery block was 16/18 correct (or 32/36 for mixed block sessions) and unprompted responses. If participants failed to meet the criteria to increase to the next prompt level, the following session was presented at the same prompt level for a second time. If a participant's performance dropped below 75% (14/18), participants were retrained beginning with 0 s delay prompt. Following the 5 s delay prompt, participants, were presented with a post mastery block, unless they met criteria to move to a block with a prompt delivered with

a shorter delay. During the post mastery block, no prompts were presented and there was no auditory feedback provided for correct responses.

Following training of two sets of relations (e.g. A_{1-3}-B_{1-3} and B_{1-3}-C_{1-3}), retraining with the first relation was conducted to ensure maintenance of matching accuracies. Retraining was achieved through the presentation of mixed block trials, defined as trials where both sets of relations were presented within the same trial, resulting in 36 trials. The same 0 s, 3 s, and 5 s progressive delay prompting hierarchy was utilized and the same level of integrity was continued during this mixed block training. Once the participant had met mastery criteria for the relations presented in a mixed format, the relations were presented in a mixed post mastery block, where again no prompts or indicators of correct responding were presented.

Once the mixed post-mastery block was completed, participants were tested for the emergence of transitivity and equivalence relations. Instructions for transitivity and equivalence blocks read as follows: "The slides you will see will each contain a sample stimulus, located at the top of the screen and three comparison stimuli, located at the bottom of the screen. These slides will not include prompts. You will not hear any sounds to indicate correct responses." Transitivity and equivalence tests consisted of a block of 18 slides in which previously mastered stimuli were presented but not in ways which had been previously trained. For example, when relations A_{1-3}-B_{1-3} and B_{1-3}-C_{1-3} were trained, transitivity probe blocks tested for the relation A_{1-3}-C_{1-3}, where A_1, A_2, and A_3 were presented as samples and C_1, C_2 and C_3 as comparisons. Equivalence probe blocks tested for the relation C_{1-3}- $A_{1-3,}$ where C_1, C_2 and C_3 presented as samples and A_1, A_2, and A_3 presented

as comparisons. Criteria for positive demonstration of equivalence and transitivity probes was determined as participants scoring 16/18 correct for each block.

Results

Table 2 represents a comparison of the number of errors and trials required for mastery across integrity values and participants. It also depicts the results of tests for both transitivity and equivalence for each participant. For two participants, Tim and Jenn, varying the degree of procedural integrity negatively affected the acquisition of stimuli learned at lowered integrity values. However, for one participant (Lisa), the integrity value did not affect the rate of acquisition or number of errors, in that the session which she demonstrated the greatest number of errors and sessions to mastery was the session taught at the highest degree of integrity (100%). Tim and Jenn required the fewest number of sessions to mastery (9 & 11, respectively) and scored the fewest number of errors (0 & 2, respectively) across both acquisition and testing phases with 100% integrity. The third participant, Lisa, however, demonstrated the highest number of errors (34) and required the most sessions to mastery (16) during the 100% integrity condition. During the 83% integrity condition, all three participants demonstrated similar numbers of errors (Lisa- 4, Tim- 3 & Jenn-3) and number of sessions to mastery (Lisa – 12, Tim – 10, Jenn -12). Finally, during the 50% integrity condition, two of the participants (Tim & Jenn), scored the highest number of errors (17 & 9, respectively). Tim required the greatest number of sessions to mastery in the 50% integrity condition (11), while Jenn demonstrated the same number of sessions to mastery (12) as seen within the 83% integrity condition.

During the 100% condition, two participants (Tim & Jenn) demonstrated both transitivity and equivalence relations among stimuli. Lisa, however, did not demonstrate

either transitivity or equivalence. During the 83% integrity condition, all participants demonstrated both transitivity and equivalence. During the 50% integrity condition, two of the participants (Lisa & Jenn) demonstrated transitivity and equivalence. Tim however did not demonstrate either transitivity or equivalence.

Figure 6 displays the percentage of prompted responses for each participant within each integrity value. Results for this graph were derived by taking into account trials where participants did not have the possibility of independently responding. Thus, for 100% integrity sessions, 100% (18) trials at the 0 s delay for the initial acquisition sessions and 100% (36) of the 0 s mixed block acquisition sessions were not used in determining the percentage of prompted responses. Within the 83% integrity condition, 83% (15) of the trials during the 0 s initial acquisition sessions and 83% (30) of the 0 s mixed block acquisition sessions were not included in the calculation of the total percentage. Within the 50% integrity condition, 50% (9) of the trials for initial acquisition sessions and 50% (18) of the 0 s mixed block acquisition sessions were not included in the total number of trials when calculating the percentage of prompted responses.

Within each integrity condition, participants showed highly varied percentages of prompted responding (50% PI range: 24%-80%, 83% PI range: 12%-81%, and 100%PI range: 10%-68%). During the 50% PI condition, Tim demonstrated the greatest percentage of prompted responses (24%), and Jenn demonstrated the second highest percentage of prompted responses (80%). However, Jenn demonstrated an almost identical distribution of responses during the 83% condition, where 81% of responses were denoted as prompted. A high level of prompted responding during both the 50% and 83% integrity condition suggests that Jenn may have developed a prompt dependence during these conditions taught at less

than 100% integrity. This effect, however, appears to have been lessened during the 100% PI condition, where she demonstrated only 68% prompted responses.

Lisa demonstrated a moderate percentage of prompted responding (51%) during the 50% PI condition, then showed an increase in the percentage of prompted responding (62%) during the 83% PI condition. During the 100% integrity condition, Lisa demonstrated the lowest percentage of prompted responding (31%).

Tim showed the highest percentage of prompted responses during the 50% integrity condition (24%), and then demonstrated a lower percentage of prompted responses during the 83% integrity condition (12%). During the 100% integrity condition, Tim demonstrated the lowest percentage of prompted responses (10%) indicating that as the level of procedural integrity increased, the percentage of prompted responses decreased.

Discussion

For Tim and Jenn, decreased levels of procedural integrity negatively affected skill acquisition, however for Lisa, decreased levels of procedural integrity appeared to have the opposite effect. Tim and Jenn both demonstrated low numbers of errors during sessions conducted with 100% integrity, and both mastered their respective relations taught with this integrity value in the fewest number of sessions. These same two participants also demonstrated a much higher number of errors, and a slightly higher number of sessions to mastery when taught their respective relations with 50% integrity. Based on these results, it can be concluded that minimal variance in the presentation time of a delayed prompt (as seen in sessions conducted at high levels of procedural integrity) cultivates the quickest and most efficient process of skill acquisition. In the same way, sessions conducted at lowered levels

of procedural integrity, where a larger variance of prompting delays occurred, appeared to hinder acquisition in two of the three participants.

Lisa's performance, however, suggests the opposite, as the highest number of errors and the greatest number of sessions to mastery occurred under the 100% integrity condition. An error analysis of her results shows that she made no errors during the acquisition of individual relations (in her case, D_{1-3}-E_{1-3} and E_{1-3}-F_{1-3}). Most errors (9) occurred within the testing phase during the post mastery mixed block trials, where both relations (D_{1-3}-E_{1-3} and E_{1-3}-F_{1-3}) were presented in the same block of trials. The remainder of her errors (25) occurred during the testing phase in both the transitivity and equivalence probes (See Figure 4). Upon further analyzing the errors that occurred during both the post mastery mixed block sessions, eight of the nine errors occurred on trials presenting the E_{1-3} to F_{1-3} relations. Analysis of the transitivity probes revealed that 100% (12/12) of Lisa's errors occurred due to the selection of F_2 in the presence of D_1 or F_1 in the presence of D_2. During the equivalence probe, 92% (12/13) of errors occurred due to the selection of D_1 in the presence of F_2 or D_2 in the presence of F_1. This analysis may lead to the conclusion that while all acquisition trials were taught at 100% integrity and no errors occurred during trials with prompting, errors that occurred during post mastery blocks with no auditory feedback present may have affected the maintenance of learned relations when they were tested during transitivity and equivalence probes. This could also lead to the conclusion that two new relations (D_1-F_2 and D_2-F_1) had been formed inadvertently during these post mastery sessions.

Another factor which may have influenced her performance is the increased amount of time elapsed between the acquisition of the D_{1-3}-E_{1-3} relations and the E_{1-3}-F_{1-3} relations

(approximately seven weeks) as compared to the other participants whose duration in time elapsed between baseline session and transitivity and extinction probe sessions ranged between one and five weeks.

All three participants demonstrated similar levels of errors (3-4) and a similar number of trials to mastery (11-12) for relations learned at 83% integrity. For two participants, Tim and Jenn, both the number of errors and the sessions required to mastery were slightly higher than those seen during their 100% integrity conditions. During the 50% integrity condition, Tim and Jenn again demonstrated similar results, making the most amounts of errors and requiring a higher number of sessions to mastery. Lisa's results during the 50% integrity condition, however, countered the other two participants, in that she demonstrated only one error (her lowest number of errors across conditions) across both acquisition and testing phases. Lisa's quick acquisition and low number of errors within the 50% integrity condition may have been partially explained due to the quick administration of her training blocks and the small lapse of time between her acquisition and testing phases (one day) as compared to the other two participants who lapsed up to seven days between completion of the acquisition phase and administration of the testing phase. This brief amount of time between training of relations could have led to better maintenance of the previously acquired relations and therefore strengthened responding.

The results explained in Table 2 may also provide insight into the emergence of transitivity and equivalence. For one participant, Lisa, neither transitivity nor equivalence were demonstrated when training was conducted with 100% integrity. Lisa also exhibited an extremely high number (9) of errors during this condition, leading to the conclusion that a substantial amount of errors during acquisition may prevent emergent relations from forming.

Tim also failed to demonstrate both transitivity and equivalence during a different integrity condition (50%). He also exhibited a substantial number of errors during acquisition (10), following the same hypothesis that a large amount of errors may preclude equivalence relations from forming. Interestingly, another participant, Jenn, demonstrated 9 errors, the same number as Lisa, during training with 50% integrity, but did, however demonstrate both equivalence and transitivity during testing.

Further research should focus on determining if an error threshold exists, that when exceeded negates the possibility of the formation of equivalence relations. Other possible research may explore if errors during training or extinction components are more detrimental to the formation of relations. Within acquisition trials, errors may be made more salient by the lack of programmed consequence (in this case, a cash register sound), however, within extinction probes, programmed consequences are not present, thus errors are unable to be observed by the responder, and may go unnoticed.

In addition to the effect which procedural integrity has on errors and number of trials to mastery, Figure 6 reveals the effect which procedural integrity has on the distribution of responding, specifically prompted versus unprompted. All participants demonstrated their lowest percentage of prompted responses and their highest percentage of independent responses during the 100% integrity condition, suggesting that while percentages of independent and prompted responding were widely varied across participants, sessions taught at 100% integrity facilitated independent responding, while sessions at lesser integrities may have hindered independent responding, and may possibly yield prompt dependency in some learners.. The large range in levels of prompted and unprompted responding also serves to demonstrate the individual nature of response distribution and the tendency (or failure) of a

student to wait for a delayed prompt. For participants such as Tim, time delayed prompts (at least those presented at a high level of integrity) appear to efficiently teach relations by minimizing errors and maximizing independence. Jenn's results, however, seem to indicate the opposite, in that, even when taught at 100% integrity (with no prompting errors) she still relied on the prompt to provide the correct response in 68% of trials.

The results of the current study suggest that procedural integrity plays an important role in the acquisition of arbitrary stimulus-stimulus relations. Decreased procedural integrity levels within this study resulted in a slower rate of acquisition of individual relations, as well as an increased number of errors. Decreased procedural integrity also produced lowered levels of independent responding for some individuals, while sessions conducted at 100% PI or as prescribed, yielded the greatest percentile of independent responding.

Investigation into whether temporal variables may influence effects of procedural integrity may warrant future research. Specifically, research conducted to further examine the effect of inter-session latency is recommended, as the current study did not explicitly measure this variable. Future research may also look at replication of this study, due to the varying results which occurred across the study's participants.

References

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

Billingsley, F., White, D. R., & Munson, R. (1980). Procedural reliability: A rationale and an example. *Behavioral Assessment, 2*, 247-256.

Cooper, J.O., Heron, T.E., & Heward, W.L. (1987). Applied behavior analysis. New Jersey: Prentice Hall.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257-263.

McIntyre, L.L, Gresham, F.M., DiGennaro, F.D. & Reed, D.D. (2007). Treatment Integrity of School-Based Interventions with Children in the Journal of Applied Behavioral Analysis 1991-2005. *Journal of Applied Behavior Analysis, 40*, 659-672.

Perry, K. (2007). Variations of Procedural Integrity and its Effects on Task Acquisition Within Chains. *Unpublished Manuscript.*

Peterson, L., Homer, A.L., & Wonderlich, S.A. (1982). The integrity of independent variables in behavioral analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.

Touchette, P.E. (1971). Transfer of stimulus control: measuring the moment of transfer. *Journal of the Experimental Analysis of Behavior*, 15, 347-354.

Vollmer, T.R., Roane, H.S., Ringdahl, J.E., & Marcus, B.A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis*, 32, 9-23.

Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis*, *39*, 369-373.

Yates, K. (2008). Varying Levels of Procedural Integrity Using a Progressive Delay Procedure to Teach Visual-Visual Stimulus Relations. *Unpublished Manuscript*.

Table 1

Effects of Varying Levels of Procedural Integrity on Errors, Trials to Mastery, and Emergence of Transitivity and Equivalence

Tim	100% procedural integrity	50% procedural integrity	83% procedural integrity
	A1-B1	D1-E1	G1-H1
	A2-B2	D2-E2	G2-H2
	A3-B3	D3-E3	G3-H3
	B1-C1	E1-F1	H1-I1
	B2-C2	E2-F2	H2-I2
	B3-C3	E3-F3	H3-I3
Lisa	83% procedural integrity	100% procedural integrity	50% procedural integrity
	A1-B1	D1-E1	G1-H1
	A2-B2	D2-E2	G2-H2
	A3-B3	D3-E3	G3-H3
	B1-C1	E1-F1	H1-I1
	B2-C2	E2-F2	H2-I2
	B3-C3	E3-F3	H3-I3
Jenn	50% procedural integrity	83% procedural integrity	100% procedural integrity
	A1-B1	D1-E1	G1-H1
	A2-B2	D2-E2	G2-H2
	A3-B3	D3-E3	G3-H3
	B1-C1	E1-F1	H1-I1
	B2-C2	E2-F2	H2-I2
	B3-C3	E3-F3	H3-I3

Table 2

Description of Session Type and Order

Types of Sessions	Example	Programmed Consequence for Correct Response	Prompt
Baseline	(A_{1-3}- B_{1-3})	No	No
Training/ Acquisition	(A_{1-3}- B_{1-3})	Yes	Yes
Post-Mastery	(A_{1-3}- B_{1-3})	No	No
Baseline	(B_{1-3}-C_{1-3})	No	No
Training/Acquisition	(B_{1-3}-C_{1-3})	Yes	No
Post-Mastery	(B_{1-3}-C_{1-3})	No	No
Mixed Block Baseline	(A_{1-3}- B_{1-3} and B_{1-3}-C_{1-3})	No	No
Mixed Block Training	(A_{1-3}- B_{1-3} and B_{1-3}-C_{1-3})	Yes	Yes
Post-Mastery	(A_{1-3}- B_{1-3} and B_{1-3}-C_{1-3})	No	No
Transitivity	(A_{1-3}-C_{1-3})	No	No
Equivalence	(C_{1-3}-A_{1-3})	No	No

Table 3

Comparison of Varying Degrees of Integrity on Number of Errors, Number of Trials to Mastery and the Formation of Transitivity and Equivalence Relations

Procedural Integrity Values/Participants		Errors	Trials to Mastery	Transitivity	Equivalence
100% I	**Lisa**	34	16	No	No
100% I	**Tim**	0	9	Yes	Yes
100% I	**Jenn**	2	11	Yes	Yes
83% I	**Lisa**	4	12	Yes	Yes
83% I	**Tim**	3	10	Yes	Yes
83% I	**Jenn**	3	12	Yes	Yes
50% I	**Lisa**	1	11	Yes	Yes
50% I	**Tim**	17	11	No	No
50% I	**Jenn**	9	12	Yes	Yes

Table 4

Lisa's Error Analysis at 100% PI

	Acquisition	**Post-Mastery Testing**	**Transitivity and Equivalence Testing**
Errors	0/34 (0%)	9/34 (26%)	25/34 (74%)

Figure captions

Figure 1. Example of a slide for the relation A_{2}-$B_{2.}$

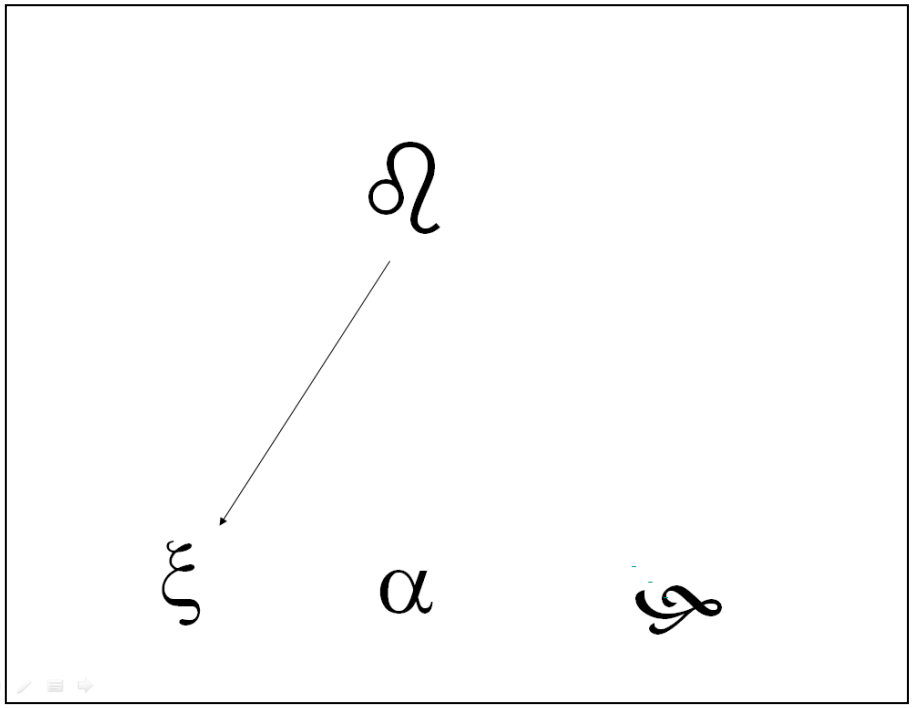
Figure 2. Stimuli $A_{1\text{-}3,}$ $B_{1\text{-}3,}$ and $C_{1\text{-}3.}$

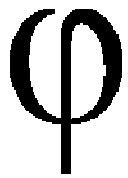
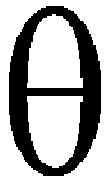
Figure 3. Stimuli $D_{1\text{-}3,}$ $E_{1\text{-}3}$, and $F_{1\text{-}3.}$

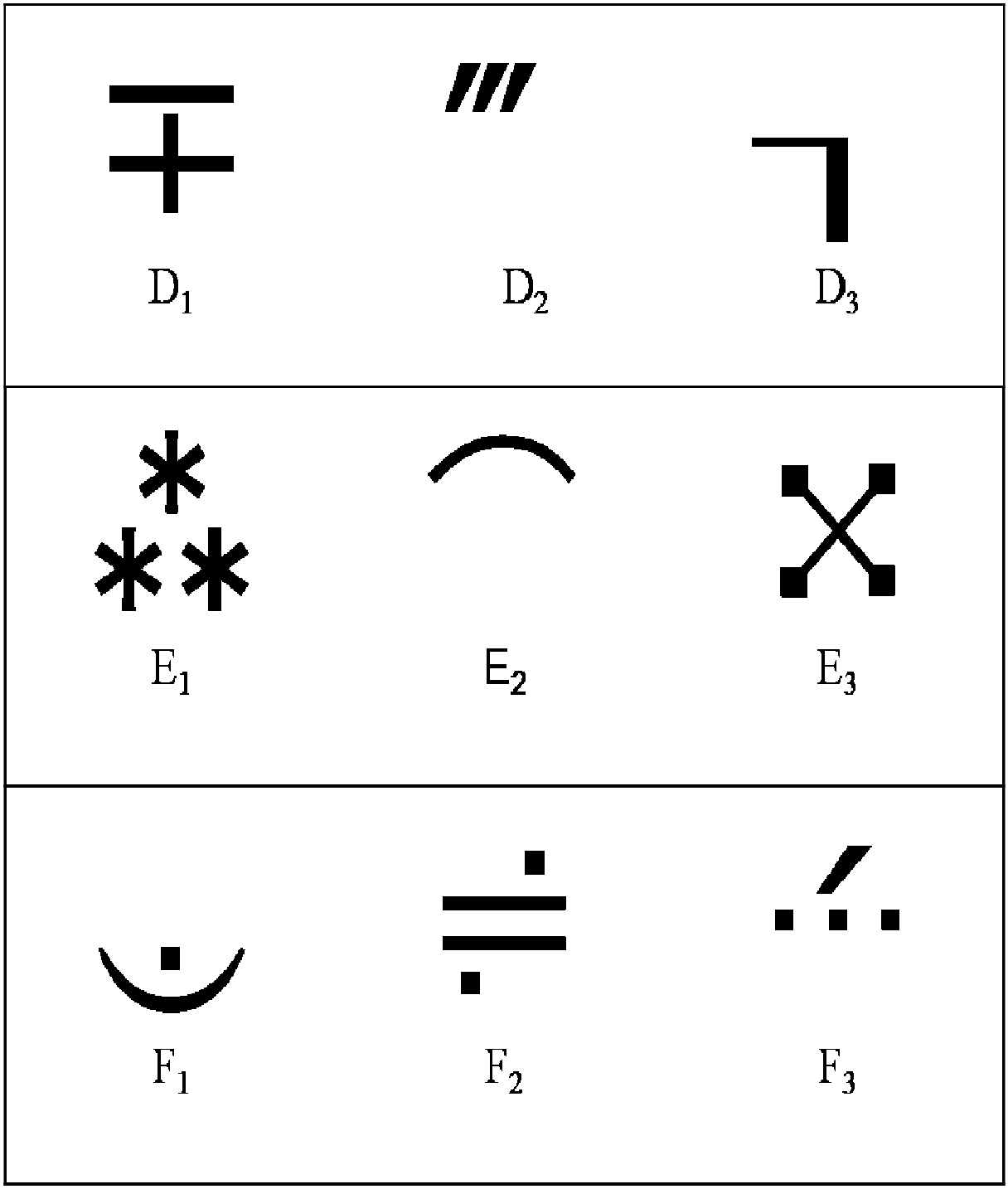
Figure 4. Stimuli $G_{1\text{-}3,}$ $H_{1\text{-}3}$, and $I_{1\text{-}3.}$

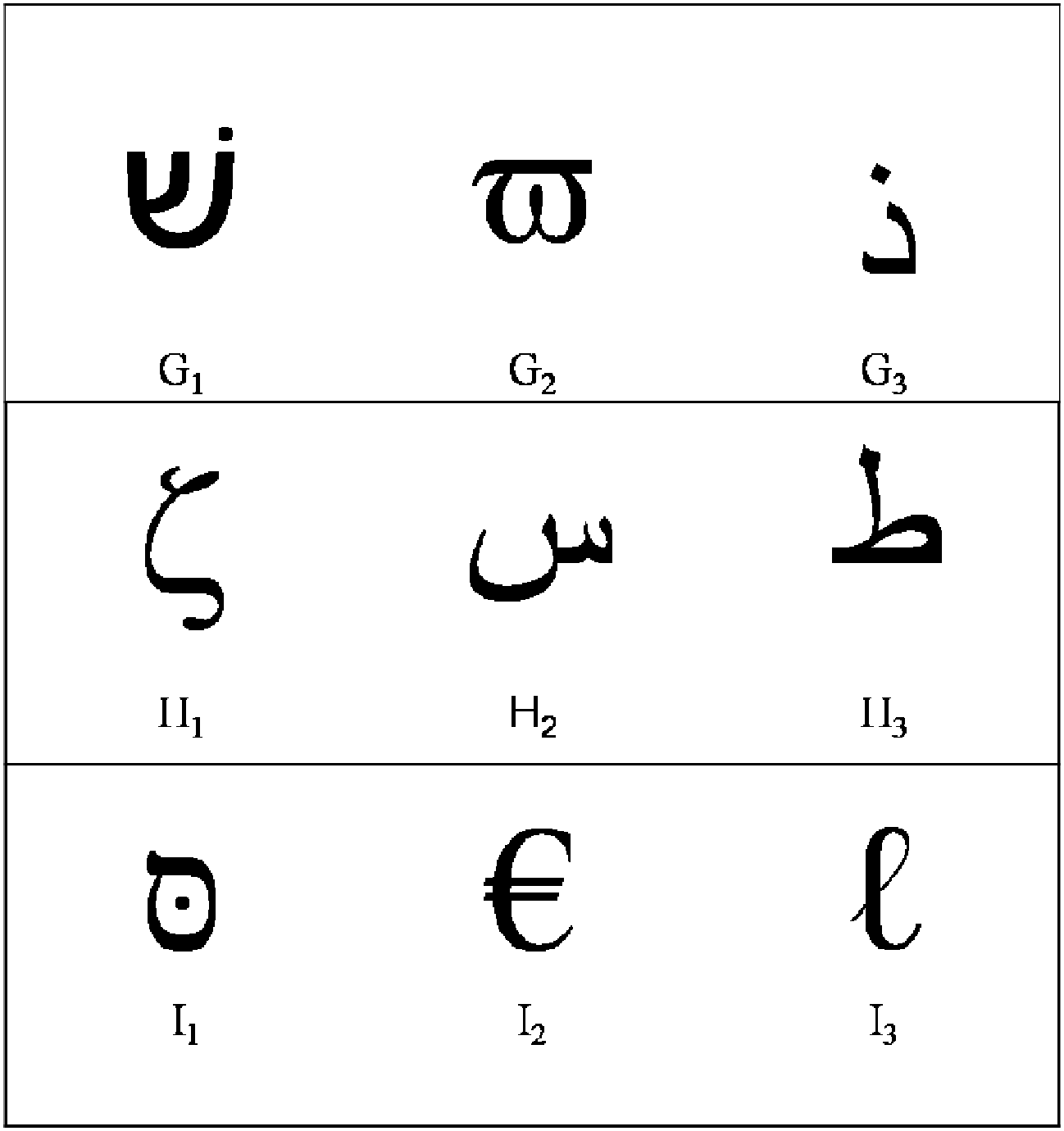
Figure 5. Data sheet for D-E relation.

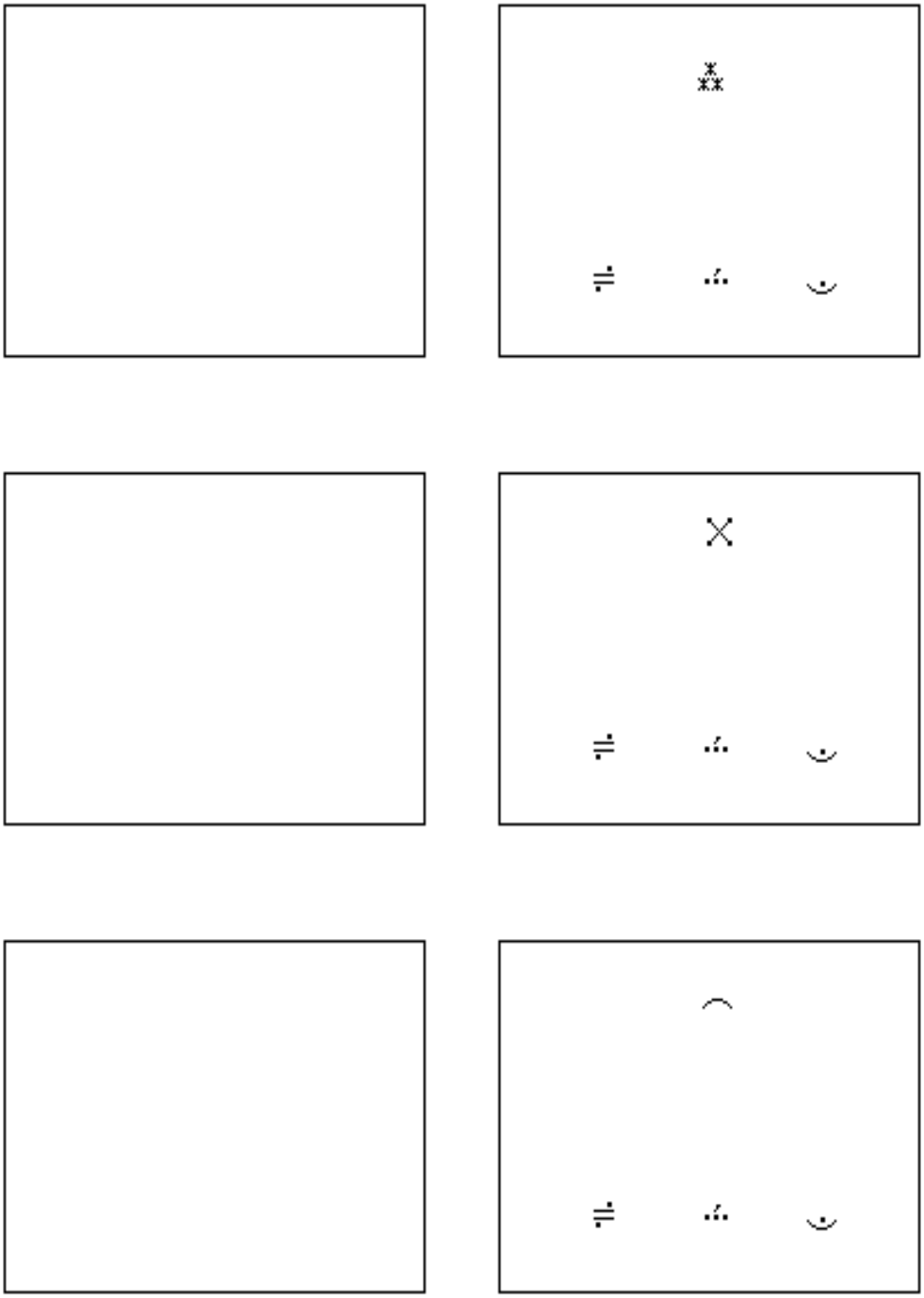
Figure 6. Percentage of prompted responses across procedural integrity values and participants.



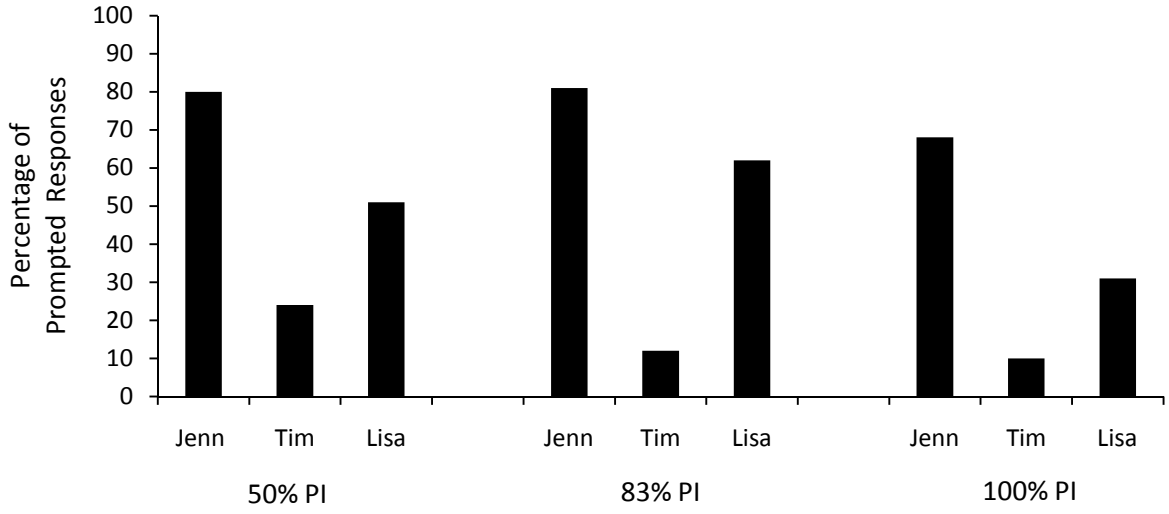
 A_1	 A_2	A_3
B_1	B_2	B_3
C_1	C_2	C_3


D1
D2
D3
E1
E2
E3
F1
F2
F3


G1
G2
G3
H1
H2
H3
I1
I2
I3



1


Percentage of Prompted Responses
100
90
80
70
60
50
40
30
20
10
0
Jenn
Tim
Lisa
Jenn
Tim
Lisa
Jenn
Tim
Lisa
50% PI
83% PI
100% PI