

AN EXAMINATION OF THE EFFECTS OF FEEDBACK ACCURACY ON ACADEMIC
TASK ACQUISITION IN ANALOGUE SETTINGS

By

Copyright 2013

Jason M. Hirst

Submitted to the graduate degree program in the Department of Applied Behavioral Science and
the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the
degree of Master of Arts.

Chairperson Florence D. DiGennaro Reed, Ph.D., BCBA-D

Derek D. Reed, Ph.D., BCBA-D

David P. Jarmolowicz, Ph.D.

Date Defended: January 22nd, 2013

The Thesis Committee for Jason M. Hirst

certifies that this is the approved version of the following thesis:

AN EXAMINATION OF THE EFFECTS OF FEEDBACK ACCURACY ON ACADEMIC
TASK ACQUISITION IN ANALOGUE SETTINGS

Florence D. DiGennaro Reed, Ph.D., BCBA-D

Date Approved: January 22nd, 2013

Abstract

Performance feedback is a common procedure used in a variety of settings to change behavior. Although reviews of the literature have identified a number of dimensions of performance feedback that are predictive of effectiveness, little research has examined the influence of inaccurate feedback on behavior. The purpose of the present study was to examine both the short- and long-term effects of inaccurate feedback on the acquisition of match-to-sample tasks. The first study adopted a translational, human operant paradigm to evaluate the effects under highly-controlled conditions. Undergraduate students were presented an arbitrary match-to-sample task on a computer. Feedback accuracy was manipulated in an initial phase followed by a condition where only accurate feedback was provided. The second study extended these findings to a more applied setting and population. The results of both studies demonstrated that exposure to inaccurate feedback resulted in the failure to acquire the tasks. Furthermore, a carryover effect was obtained represented by a delay to acquisition following the improvement of feedback accuracy. The behavioral processes behind the results obtained are interpreted through a synthesis of literatures on performance feedback, fidelity, and instructional control. Implications for educational and organizational settings are suggested.

Keywords: *performance feedback, feedback accuracy, instruction, match-to-sample, rule-governed behavior*

Acknowledgements

First and foremost, I'd like to thank Florence DiGennaro Reed for the enormous amount of support. Her mentorship, patience, and what I can only assume to be accurate feedback made this project possible. Flo's model of mentorship was invaluable to my education thus far. Being a part of a strong research team helped to build the necessary competencies while keeping research what it should be: fun! This may have been the reason that some unknown group of people nominated her for an award... Did we ever figure out who that was?

I'd also like to thank Derek Reed for his support and input on the pilot project for this study. His guidance on the basic literature made a nearly impossible task ("Cite more basic literature on learning," Reviewer 2, personal communication) more manageable. I also learned a lot about quantitative analyses, but more importantly, I learned how much I still don't know.

Thanks go out to my undergraduate research assistant for going through the 10 hours of video for IOA and fidelity. I can only imagine how tedious that must have been.

I'd like to thank the staff and the director of the pre-school research site. They were very flexible and accommodating of a research study that ended up lasting so much longer than I anticipated.

I need to thank my research team. Specifically, thank you for not judging me for wallpapering my half of the office with data and staring at it interminably. Also, thank you for taking an interest, or at least pretending to, in this work. It's hard to not talk about something that owned over a year of my life and you were still listening at the end. Also, thank you for keeping me motivated. When the deadline for this document was still months in the future, that countdown kept me from procrastinating [too much].

Thanks are deserved for coffee, single-serving coffee machines, microwavable meals, and internet radio. Just think how much longer this would have taken if I actually had to get up and go to lunch every day and sleep for a reasonable amount of time every night...

Finally, I want to thank my friends and family who, despite not having any idea what it is I do all day in my office and research sites, persistently provided me with additional support. Thanks for being there through the highs and lows.

Also, thank everyone whom I am missing here. Make sure you mark off that space on your seminar bingo card!

Table of Contents

Introduction	1
Study 1	
Methods	20
Results and Discussion	28
Study 2	
Methods	34
Results and Discussion	40
General Discussion	44
References	56
Tables	66
Figures	72
Appendices	90
Vita	120

List of Illustrative Materials

Table 1.	Target shapes and corresponding names	66
Table 2.	Significance levels of Dunn’s Multiple Comparison Test. Data above the diagonal represent significance levels of differences between groups in the inaccurate feedback condition. Data below the diagonal represent significance levels between groups during the accurate feedback condition. Non-significant differences are marked as ns.	67
Table 3.	List of instructions delivered for each task. For instructions not specifying a specific response topography, any discrimination response was accepted as a response.	68
Table 4.	Interobserver agreement by participant and condition.	69
Table 5.	Procedural fidelity by participant and condition.	70
Table 6.	Preferred items identified during the preference assessment for each participant	71
Figure Captions		72
<i>Figure 1.</i>	Screenshot of the computer interface during the match-to-sample task displaying “Correct” feedback in green below the sample stimulus.	74
<i>Figure 2.</i>	Screenshot of the computer interface during the match-to-sample task displaying “Incorrect” feedback in red below the sample stimulus.	75
<i>Figure 3.</i>	Cumulative records for participants in the 25% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.	76
<i>Figure 4.</i>	Cumulative records for participants in the 50% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.	77
<i>Figure 5.</i>	Cumulative records for participants in the 75% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.	78

<i>Figure 6.</i>	Number of trials before mastery criterion of 15 consecutive correct responses was met displayed by group. Each data point represents one participant. Data points below the delta on the Y-axis, in the shaded region, represent mastery before the phase change from inaccurate feedback to accurate feedback.	79
<i>Figure 7.</i>	Percent correct responses during the inaccurate feedback condition by group. Each data point represents one participant. The dotted horizontal line at 0.2 represents chance level responding. The solid horizontal lines display the arithmetic mean for each group.	80
<i>Figure 8.</i>	Percent correct responses during the accurate feedback condition by group. Each data point represents one participant. The dotted horizontal line at 0.2 represents chance level responding. The solid horizontal lines display the arithmetic mean for each group.	81
<i>Figure 9.</i>	Standardized percent area under the curve during the inaccurate feedback condition by group. Solid horizontal lines are the arithmetic mean for each group.	82
<i>Figure 10.</i>	Standardized percent area under the curve during the accurate feedback condition by group. Solid horizontal lines are the arithmetic mean for each group.	83
<i>Figure 11.</i>	Index of strength of incorrect discriminations as represented by the proportion of incorrect selections of comparison stimuli in the presence of each sample stimulus for all participants by group.	84
<i>Figure 12.</i>	Data for participant A.	85
<i>Figure 13.</i>	Data for participant B.	86
<i>Figure 14.</i>	Data for participant C.	87
<i>Figure 15.</i>	Data for participant D.	88
<i>Figure 16.</i>	Index of strength of incorrect discriminations as represented by the proportion of incorrect selections of comparison stimuli in the presence of each sample stimulus for all participants by condition (top panel) and by task (bottom panel).	89

Appendix Index

A) Participant recruitment script	90
B) Demographic survey	91
C) Participant introduction script	92
D) Indices of discrimination strength	93
E) Parent letter	109
F) Informed consent form	110
G) Assent procedure	114
H) Insect task stimuli	115
I) River task stimuli	116
J) Country task stimuli	117
K) Moon task stimuli	118
L) Sample data sheet	119

An Examination of the Effects of Feedback Accuracy on Academic Task Acquisition in Analogue Settings

Practitioners are tasked with the selection and implementation of treatments that represent the best treatment for their clients based on the available evidence of effectiveness and their clients' individual needs. In today's age of escalating accountability and advocacy for evidence-based practice, empirical support for the effectiveness of treatments is becoming an area of expanding interest. The tie between research and practice has been a common goal among scientists for at least the past century (DiGennaro Reed & Reed, 2008). However, the current trend of interest in evidence-based practice is built on a relatively recent movement with beginnings in the United Kingdom under the title of evidence-based medicine (Chambless & Ollendick, 2001).

Sackett and colleagues (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996) define evidence-based medicine as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (p.71). The authors propose that neither clinical expertise nor empirical evidence from research are sufficient alone, but that physicians should integrate the two. Several challenges to this approach have been identified. Physicians have reported that they do not have enough time to devote to reviewing newly emerging research. A self-report survey of medical consultants in the United Kingdom revealed that some consultants devote less than one hour per week to keeping abreast of the literature (Sackett et al., 1996). Under these conditions, it is likely that the services medical practitioners perform will have less correspondence to "best practices" as time progresses since their training. Evidence from the literature has emerged suggesting that deviations from prescribed procedures during the implementation of an empirically-supported treatment may result in divergent or

differential outcomes. Although the literature is growing in scope, the type and effect of implementation errors is vast and to date still not well understood.

The resonance of these issues in other fields is evident in the activities of professional organizations and the U.S. government. Within psychology, evidence-based practice has received significant attention. In 1993, Division 12 (clinical psychology) of the American Psychological Association (APA) adopted a report by the Task Force on Promotion and Dissemination of Psychological Procedures, which was the first attempt to standardize criteria within the field of clinical psychology for describing a treatment as empirically-validated or empirically-supported (APA, 1995). In order to be considered a well-established, empirically-validated treatment, effectiveness must be demonstrated in at least two between-group designs, the evidence needs to support the effectiveness of the treatment against either a control group or another validated treatment, the treatment must be manualized, the characteristics of the sample must be clearly specified, and the effects of the study must be demonstrated across two researchers or research teams. Criteria for single-case designs are similar with the number of studies demonstrating effectiveness being more than nine (Chambless et al., 1998). The task force further committed Division 12 to maintaining and updating a list of treatments meeting criteria for the benefit of practicing psychologists who may not otherwise have the resources to incorporate the most recent and best research into practice.

The commitment to maintaining a resource for practitioners is embodied by the What Works Clearinghouse (WWC), a service maintained by the Institute of Education Sciences, U.S. Department of Education. The purpose of the WWC is to provide a credible and centralized source of empirically-supported practices in education to educators, who frequently do not have the time or resources to devote to reading relevant literature (What Works Clearinghouse, n.d.).

The standards for evidence used by WWC are strongly designed around randomized controlled trials, although the most recent version of the Procedures and Standards Handbook contains pilot criteria for evaluating both quasi-experimental and single-case designs.

Concurrently with the development of standards for describing what constitutes evidence, governing bodies have begun to incorporate an imperative for practitioners to incorporate the use of scientific evidence into their work. For example, standard 2.04 of the APA Ethical Principles of Psychologists and Code of Conduct states, “Psychologists’ work is based upon established scientific and professional knowledge of the discipline” (APA, n.d., p. 5). The Behavior Analyst Certification Board Guidelines for Responsible Conduct requires that recommendations are based on empirical evidence of both long- and short-term effectiveness (BACB, n.d., Section 2.09a). In another example, the federal government passed the Individuals with Disabilities Education Improvement Act (2004) and the No Child Left Behind Act (2001) mandating the use of evidence. Taken together, these standards have fostered greater accountability amongst practitioners to use interventions that work, and are in the best interest of those receiving services.

The push for standards of evidence and corresponding mandates for their use are not without flaws. The guidelines described above are substantially biased towards well-controlled studies of efficacy with criteria typically met by randomized controlled trials, at the exclusion of effectiveness research conducted in more naturalistic, less controlled settings. Additionally, some criteria do not include an explicit requirement for the monitoring of the fidelity of the independent variable (e.g., APA, cf. Chambless & Hollon, 1998). However, calls for the inclusion of fidelity measurement in standards of evidence-based practice have been made recently (e.g., Detrich, 2008; DiGennaro Reed & Reed, 2008; Horner, Carr, Halle, McGee,

Odom, & Wolery, 2005). More recent standards developed in the fields of education and autism treatment have brought the degree of implementation of the independent variable into consideration. For example, the pilot standards for single case research studies for the WWC include measurement and quantification of implementation of the independent variable (WWC, n.d.). The National Autism Center's National Standards Project includes requirements for measurement of independent variable implementation in both the Scientific Merit Rating Scale and the Treatment Effect Rating components of their evidence standards (National Autism Center, n.d.).

Interventions with strongly demonstrated efficacy during randomized controlled trials have often resulted in less significant results when implemented in real world settings (Hoagwood, Hibbs, Brent, & Jensen, 1995). This discrepancy between efficacy and effectiveness is typically attributed to factors of external validity, such as differences between research participants and recipients of the services, and less control over extra-experimental variables (Boruch & Gomez, 1977). Less considered is the degree to which the implementation of the treatment variables in therapeutic contexts (i.e., procedural fidelity) matches the implementation during examinations of efficacy.

Procedural Fidelity

Procedural fidelity, or treatment integrity, has become a topic of discussion in many disciplines, especially the human service fields such as community-based intervention (Breitenstein et al., 2010), clinical psychology (e.g., Perepletchikova, 2011), nursing (Song, Happ, & Sandelowski, 2010), learning disabilities (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000), special education (Duhon, Mesmer, Gregerson, & Witt, 2009), emotional and behavioral disorders (Lane, Jolivette, Conroy, Nelson, & Brenner, 2011), and behavior analysis

(Peterson, Homer, & Wonderlich, 1982). Collectively, the broader literature emphasizes the critical nature of implementation monitoring, in all service disciplines. Each of the aforementioned fields has had its own call for the study of and devotion of resources to fidelity measurement and improvement. The impetus for these calls vary by field but share a common theme; the measurement and reporting of fidelity data is required in order to make informed decisions and to draw valid conclusions about treatment effects. Beyond the need to report the data, researchers from multiple fields have focused their efforts on defining and developing practical ways of measuring fidelity (e.g., Song et al., 2010). Much work still remains to be done.

The literature has used several terms to describe factors related to independent variable or treatment implementation such as treatment integrity (Yeaton & Sechrest, 1981), fidelity (Moncher & Prinz, 1991), adherence (Dane & Schneider, 1998), and procedural fidelity (Wolery, 1994). These terms have been used interchangeably, or used with varying degrees of specificity regarding what is being measured. Yeaton and Sechrest (1981) defined treatment integrity as the degree to which a treatment is implemented as intended, although this definition has been unpacked and expanded over time. More recent literature has suggested that relatively simple conceptualizations and definitions of treatment integrity may be inadequate.

Across fields, several common dimensions of integrity have been proposed. The first and most commonly measured (Schulte, Easton, & Parker, 2009) is adherence, which has been defined broadly as the ratio of intervention steps implemented to steps prescribed (Dane & Schneider, 1998). Waltz, Addis, Koerner, and Jacobson (1993) extended adherence by further analyzing the interventionist's behavior of implementation into (a) essential and unique behaviors, or those that are both necessary and unique to the treatment being implemented, (b) essential but not unique behaviors, or those that are part of the treatment protocol but are also

shared by other treatments, (c) unnecessary but not proscribed behaviors, which are steps implemented by the interventionist that are not formally part of the intervention procedure but that are not explicitly proscribed, and (d) proscribed behaviors. The latter have also been referred to as integrity errors of commission: the addition of steps to a treatment procedure that should not be added (e.g., Vollmer, Roane, Ringdahl, & Marcus, 1999). Implementation quality is a measure of how well the steps in an intervention are implemented (Sanetti & Kratochwill, 2009). Quality has been used to describe interventionist behaviors that are not required as part of adherence, but that may influence the effectiveness of the intervention. Some conceptualizations have also referred to as interventionist competence in implementing the treatment (Waltz et al., 1993). A third dimension that has been proposed across several fields is exposure to treatment or dosage, which refers to the duration of the treatment or how much of the treatment is delivered (Dane & Schneider, 1998). Treatment differentiation refers to the degree to which the treatment being implemented or evaluated is qualitatively different than the absence of the treatment or some other treatment (Moncher & Prinz, 1991). Finally, Dane and Schneider (1998) define participant responsiveness as the degree to which the participant or client is engaged in the intervention.

Based on a distillation of these proposed dimensions of integrity, Sanetti and Kratochwill (2009) proposed an update to the definition of treatment integrity as “the extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention” (p. 448). The above can be generally applied to questions regarding the implementation of a treatment, but implementation has a broader scope in many fields. That is, within a consultation model, implementation measures can be separated into measures of consultant behaviors and consultee behaviors, or into implementation of

training and implementation of treatment (e.g., Noell, 2008). It is proposed here that the term *procedural fidelity* can be used to refer to the broadest scope of considerations for implementation.

These dimensions represent some common aspects of fidelity that are shared between conceptualizations across fields, but are not a comprehensive list of dimensions that individual disciplines have proposed. It is important to note that although the aforementioned dimensions have a high degree of face validity and conceptual relevance as metrics of treatment implementation, empirical support for the psychometric properties of treatment integrity as a multidimensional construct is still emerging (Sanetti & Kratochwill, 2009). The disagreement on a conceptual model of treatment integrity has also impaired the development of consistent methods of measuring implementation (Gearing et al., 2011). An investigation by Sanetti and Fallon (2011) examined how the application of different measurement techniques to the same intervention data yield both different results and may result in different decision-making regarding the effectiveness of an intervention. For example, the same level of implementation may yield different decisions about the cause of an effect or non-effect depending on the availability of more than one dimension of fidelity data and whether the data are analyzed by session or by intervention component.

The influence of procedural fidelity on decision-making is equally concerning in both research and practice. Boruch and Gomez (1977) proposed that a general assumption of experimental control—that treatment variables are either present or absent—is inaccurate and insufficient. The implementation of independent variables may instead occur along a continuum. The authors provide several examples of discrepant results between efficacy and effectiveness studies that can be attributed to differences in how the treatment was delivered. The push for

clearly specified and manualized treatments fostered by the evidence-based practice movement is aimed at providing clear description of procedures such that other researchers and interventionists can precisely replicate treatments (e.g., Chambless et al., 1998). However, consumers of research and interventions should not rely on the assumption that the treatment was implemented precisely as described. Rather, conclusions about functional relations between independent and dependent variables in research should be withheld unless independent variable reliability is monitored and reported, and confidence is high that the intervention was implemented accurately and consistently (Peterson et al., 1982). In applied practice, the application of evidence-based practice demands similar considerations. The evaluation of treatments in the absence of information on the fidelity of implementation may lead to inaccurate or faulty decisions about client responsiveness to intervention and may influence treatment decisions to continue, terminate, or modify interventions. If a client fails to respond to a treatment, a more restrictive treatment may be imposed inappropriately, which could violate a client's right to the most effective and least restrictive treatment (Van Houten et al., 1988). In special education settings, decisions on eligibility for services are made based on response to a pre-referral intervention, and if the intervention is implemented incorrectly, students may be inappropriately referred to special education (e.g., Duhon et al., 2009).

Despite the presence of calls for fidelity measurement in the literature (e.g., Billingsley, White, & Munson, 1980; McIntyre, Gresham, DiGennaro, & Reed, 2007), Peterson et al. (1982) describe a "curious double standard" that has developed, which refers to a disproportionate focus on monitoring the reliability of dependent measures but not independent variables (p. 478). Several literature reviews have illustrated this persistent trend (e.g., Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; McIntyre et al., 2007). A survey of researchers in clinical

psychology indicated that barriers to the monitoring and reporting of fidelity measures in research do not include a perceived lack of importance, but rather a lack of guidelines and a theoretical framework for fidelity (Perepletchikova, Hilt, Chereji, & Kazdin, 2009). Sanetti and DiGennaro Reed (2012) extended these findings to researchers in school psychology with similar results. Some researchers have begun to solidify a comprehensive method for monitoring fidelity (e.g., Gearing et al., 2011), but a growing body of literature has also illustrated the usefulness of even rudimentary metrics of fidelity in research and applied settings.

An analysis of published studies on school-based behavioral interventions between 1980 and 1990 correlated treatment integrity with effect size ($r = .51$) and percentage of non-overlapping data ($r = .58$) (Gresham et al., 1993). Other studies using treatment integrity as a dependent variable noted improvements in treatment outcomes concurrently with improvements in integrity. In two studies, DiGennaro and colleagues (DiGennaro, Martens, & Kleinmann, 2007; DiGennaro, Martens, & McIntyre, 2005) trained teachers to implement school-based interventions for problem behaviors among students. As the integrity with which the teacher participants implemented procedures improved, a concomitant decrease in student problem behavior was observed. A post-hoc analysis showed that treatment integrity and treatment effectiveness were significantly correlated. In a similar study, Dib and Sturmey (2007) used modeling, feedback, and rehearsal to improve the integrity with which three teaching aides implemented a discrete-trial instructional procedure. Concurrently with improvements in the implementation of the procedure, a significant decrease was observed in student engagement in stereotypic behaviors during instruction. Studies such as these have provided at least correlational evidence of a relationship between integrity and outcome.

Further research experimentally manipulated treatment integrity in order to provide causal evidence of the relation with measures of outcome. Several studies have assessed the impact of degradations in treatment integrity on the maintenance of treatment effects following perfect implementation. Northup, Fisher, Kahng, Harrell, and Kurtz (1997) assessed the necessary level of treatment integrity, referred to as treatment strength, to maintain the effects of a differential reinforcement of alternative behavior (DRA) and time-out procedure. Reinforcement was available on a variable interval schedule for appropriate responses targeted in the DRA. The level of implementation was manipulated by doubling the interval schedule value to represent 50% strength and doubling again to represent 25% strength. The integrity of the time-out procedure was manipulated by only implementing time out on variable ratio 2 or variable ratio 4 schedules representing 50% and 25% integrity, respectively. The results showed that for two of three participants, treatment effects were maintained when both procedures were implemented at 50% of initial integrity. For the third participant, treatment effects were maintained when time-out was implemented at 25% integrity.

In related studies (Vollmer, Roane, Ringdahl, & Marcus, 1999; St. Peter Pipkin, Vollmer, & Sloman, 2010), the integrity of a DRA procedure was manipulated by changing the integrity of the reinforcement and extinction components separately. Treatment integrity was measured as the proportion of appropriate responses that were reinforced or the proportion of inappropriate responses that were not reinforced during extinction. In both studies, the results indicated that following the perfect implementation of the DRA, treatment effects were maintained until the schedule of reinforcement for inappropriate behavior was richer than the schedule for appropriate behavior. That is, behavior was consistently allocated to the response receiving more frequent reinforcement.

Taken together, these studies indicate a relatively consistent influence of the integrity of reinforcement procedures on the effectiveness of procedures designed to reduce the incidence of problem behaviors. A casual observation of the conditions representing different levels of integrity supports the attribution of the effects to differing schedules of reinforcement for both appropriate and inappropriate responses. That is, imperfect integrity may not reduce the effectiveness of the procedures, but effectively bring behavior under the control of qualitatively different reinforcement schedules.

A series of studies examined the effects of the integrity with which antecedent instructional procedures were implemented on skill acquisition, but in the presence of initially low integrity. Wilder, Atwell, and Wine (2006) manipulated the integrity of a three-step prompting procedure. Three instructions with which participants did not comply were assigned to one of three levels of integrity: 0%, 50%, or 100%. Integrity was manipulated by changing the proportion of trials where the antecedents were delivered, which corresponded to the level of integrity. The results indicated that participants did not comply with instructions associated with the 0% integrity condition. Compliance increased significantly when the prompting procedure was implemented during all trials and the 50% integrity condition resulted in levels of compliance between the other conditions.

Grow et al. (2009) also examined integrity errors within a prompting procedure. Two tasks were taught to each participant using either a system of least prompts—the prescribed procedure—or multiple verbal prompts in a multielement design. The use of multiple verbal prompts represented an integrity error by committing extra verbal prompts and omitting the more intrusive prompts associated with the system of least prompts. Tasks associated with the system of least prompts were consistently mastered before the tasks associated with multiple verbal

prompts only. Although the difference between rates of mastery was small, the authors propose that over a period of time, a compounding effect of integrity errors may significantly delay learning. That is, if each task requires more instructional time to master, over time, the student will not have acquired the number of skills as may have been possible if teachers had adhered to the validated, effective procedure.

To this point, the research summarized on the effects of integrity errors on learning has focused on antecedent procedures. However, the provision of reinforcement is also an important component of establishing and maintaining academic behaviors. To examine the influence of commission errors—adding steps not prescribed to a procedure—of consequence-based procedures on learning, DiGennaro Reed, Reed, Baez, and Maguire (2011) used a discrete trial instruction procedure to teach an arbitrary match-to-sample task to students with developmental disabilities. When no errors were made, participants showed the best performance. However, the results were more variable when some or many errors were committed. For one participant, a clear difference was observed between the 50% errors condition, where every other error was reinforced, and the 100% errors condition, where every error was reinforced. The participant responded correctly during approximately 50% of trials in the 50% errors condition and around 10% during the 100% errors condition. For the other two participants, performance was undifferentiated between the two conditions during which errors were committed.

Although much of the research on fidelity supports the general conclusion that the best outcomes are achieved under the highest levels of fidelity, there is still a need for additional research to address gaps and discrepancies. For example, consistent results are not obtained when the fidelity of antecedent procedures are manipulated compared to consequence-based procedures. There is significant variation in the magnitude and nature of the influence of

imperfect fidelity across studies, and across participants within studies. Also, given that fidelity may refer to the implementation of any number of intervention steps or techniques, further research is necessary to examine the effects of fidelity on the effectiveness of more procedures including those commonly used in organizational and educational settings, such as performance feedback.

Performance Feedback and Rule-Governed Behavior

Performance feedback is perhaps the most commonly used procedure within organizational behavior management, with more than half of the reports in the *Journal of Organizational Behavior Management* including some form of feedback in their methodology (Alvero, Bucklin, & Austin, 2001). Feedback also plays an important role in education. That is, student performance is likely influenced, at least in part, by feedback from teachers in the form of grades, verbal feedback, or written feedback. Recent studies have evaluated the use of performance feedback to address performance deficits in several academic skills with positive effects (Eckert, Dunn, & Ardoin, 2006).

Although performance feedback has commonly appeared as a procedure in behavior analytic journals and is often used in practice, it is still not well understood. Feedback is most commonly defined as information delivered to an individual regarding past performance (Prue & Fairbank, 1981). Daniels (1994) extended this definition to include a requirement that the information also indicate how the individual can improve his/her performance in the future.

Feedback as a term can describe a vast array of procedures and methods (Ford, 1980). Mirroring discussion over the format and dimensions of effective feedback is the debate over what behavioral function it serves. Peterson (1982) argues that feedback serves any number of functions. The hypothesized functions of feedback fall into two general categories: reinforcement

and stimulus control (Duncan & Bruwelheide, 1986). A reinforcement function of feedback would be inferred if feedback is delivered following a performance and performance improves as a result (e.g., Prue & Fairbank, 1981). The literature contains some examples of feedback appearing to serve as a reinforcer (i.e., feedback increases performance and/or desired behaviors; e.g., Komaki, Barwick, & Scott, 1978). Feedback may serve as a conditioned reinforcer if it is reliably associated with some other source of reinforcement (Peterson, 1982). However, feedback is not likely to only serve a reinforcement function as there are instances where feedback is delayed significantly after a performance, which reduces the chance that reinforcement has played a role. A second possibility is that feedback serves a discriminative function. That is, feedback is often provided prior to the next opportunity for a response, and may serve to prompt and evoke the behavior that will be reinforced (e.g., Daniels, 1994). If feedback is provided immediately prior to an opportunity to respond, and effectively evokes the desired behavior, it could be said that it functioned as a discriminative stimulus. Aside from a discriminatory function, a second antecedent function that feedback may serve is as an establishing operation. Duncan and Bruwelheide (1986) propose that feedback may alter the value of behaviors and outcomes by providing a description of a relation to obtaining other reinforcers.

Feedback may potentially serve any or all of the aforementioned functions, depending on the context. However, Agnew and Redmon (1993) argue that feedback does not consistently fit the definitions of either reinforcement or discriminative stimuli. They propose, instead that feedback constitutes a rule, or contingency-specifying stimulus (CSS). Skinner (1969) first described a class of operant behavior referred to as rule-governed behavior. He distinguishes rule-governed behavior from contingency-shaped behavior in that the latter requires an organism

to come into contact with the contingencies in the environment before behavior can be changed, while the former requires only that a verbal description of the complete contingency (i.e., antecedents, behavior, and consequences) be delivered. Although CSSs are typically defined as having a discriminative function, it is not always the case that CSSs immediately evoke behavior, but rather they serve to alter the function of other discriminative stimuli, establishing operations, reinforcers, and punishers (Schlinger & Blakely, 1987). This function-altering effect of CSSs facilitates the explanation of more complex human behaviors where behavior occurs neither immediately following, nor immediately preceding feedback. Among these complex behaviors is academic response acquisition.

Given the frequency of the use of feedback in applied settings, the literature has examined the necessary and important dimensions and components of feedback. For example, a review by Alvero and colleagues (2001) examined the literature for characteristics of feedback associated with the best and most consistent effects. They examined feedback alone and in combination with other procedures, as well as the dimensions of feedback including the source, format, frequency, content, and privacy. Perhaps implied is that effective feedback is accurate (i.e., delivered as intended). However, in practice, this may not always be the case.

There is limited research on the influence of inaccurate feedback on socially-relevant behaviors, but the basic literature has provided a foundation on which future research might build. In several basic research studies the degree to which instructional control influenced behavior during either contingencies of reinforcement that corresponded with instructions (matched) or contingencies that did not correspond with instructions (unmatched) was examined (e.g., Kaufman, Baron, & Kopp, 1966; Galizio, 1979). Kaufman et al. (1966) provided instructions to participants before an operant task. All participants were placed on a variable-

interval schedule of monetary reinforcement, but some participants were given inaccurate instructions indicating that the schedule in place was either a fixed-interval or variable-ratio schedule (i.e., unmatched contingencies). Participants given inaccurate instructions responded as if the schedule in place was the one instructed rather than the actual schedule, suggesting that instructional control may override or effectively compete with schedule-controlled behavior. Galizio (1979) extended these findings by studying whether a large discrepancy between instructions and actual contingencies of reinforcement would reduce instruction-following. Participants were placed in a room with a lever and given instructions to pull the lever on a certain schedule in order to avoid a monetary loss. When the instructions provided were accurate (i.e., matched contingencies), participants consistently followed the instructions and avoided the loss of monetary reinforcers. However, when inaccurate instructions resulted in the participants contacting the monetary loss, participant responding deviated from the instructions in favor of response patterns minimizing loss. In sum, these studies demonstrated that in the presence of instructions that inaccurately describe underlying reinforcement contingencies, participants reliably followed rules despite rule-following failing to maximize the rate of reinforcer delivery. However, contact with the underlying contingencies (especially response cost contingencies) effectively competes with instructional control.

Relatedly, DeGrandpre and Buskist (1991) showed that when participants were able to contact discrepancies between instructions and reinforcement contingencies, rule-following was related to the degree of accuracy of instructions. In their experimental preparation, participants were asked to press either the 'A' key or 'B' key on a keyboard and were given instructions on which to press. When the instructions were accurate, following the instructions resulted in earning points. When the instructions were inaccurate, following the instructions did not earn

points, but responding counter to the instructions resulted in points earned. The results revealed that when instructions were completely accurate, participants consistently followed the rules stated. When instructions were completely inaccurate, participants consistently deviated from the instructions. Finally, when instructions were 50% accurate, rule-following was variable. The authors propose that the results are related to the correspondence of rule-following with reinforcement. That is, when instructions are accurate, rule-following is differentially reinforced, while, in the presence of inaccurate instructions, behavior that deviates from the rule is differentially reinforced.

Ribes and Rodriguez (2001) examined the degree to which inaccurate instructions influenced both performance on a match-to-sample task and self-descriptions of the contingencies of reinforcement. The experimental preparation involved a conditional discrimination task in which participants were presented with sample stimuli and comparison stimuli. Both similarity matching (choosing the comparison stimulus that resembles the sample) and difference matching (choosing the comparison that differs from the sample) were tested. Participants were either given correct instructions—instructing them to choose the similar stimulus for similarity matching and the dissimilar stimulus for difference matching—or incorrect instructions for which the instructions were reversed. Following each trial, participants were asked to choose from a list of seven textual statements that described their performance on that trial. For example, during similarity matching, a participant choosing a correct self-description of his/her performance would respond that they chose the comparison stimulus because it was similar. The results showed that when instructions were accurate, participants were able to emit correct responses to the task and were also able to accurately describe the contingencies. If inaccurate instructions were given, participants emitted more inaccurate

descriptions of the actual contingencies, but as they contacted the contingencies, both performance and descriptions of the contingencies became more accurate.

Taken together, these studies suggest that when instructions are provided prior to responses, rule-governed behavior can compete with response patterns expected according to the underlying contingencies. However, when contact is made with underlying contingencies, contingency-shaped behavior prevails. In the absence of contact with contingencies of reinforcement or punishment, however, inefficient responding persists. DiGennaro Reed, Miller, Hirst, Reed, and Kaplan (under review) examined the framing of instructions and its influence on responding. Participants received instructions regarding the optimal pattern of responding that initially accurately described the programmed contingencies. Throughout the study, the contingencies changed such that the instructions were no longer accurate. Participants who were told they “must” follow the instructions consistently adhered to the instructions even after contacting discrepant contingencies of reinforcement. Participants who were only told to “consider” following the instructions changed their performance to match the contingencies, rather than the instructions.

The aforementioned studies have focused on rule-governed behavior established by instructions. Few studies have extended this line of inquiry to rule-following established by feedback provided following responses. Contact with contingencies for effective and ineffective behavior may result in the construction of verbal descriptions of the contingencies by verbal humans, which may function as rules for how to respond in the future (Ribes, 2000). Similarly, feedback regarding a previously made response may serve as a CSS for future responses. That is, if feedback is received that a response is correct, a contingency is specified that in the same context, the same response will be reinforced. If feedback is received that a response was

incorrect, the next presentation of the context should serve as an S^Δ for the same response. If, however, the feedback is inaccurate, subjects may develop inaccurate rules regarding which response will result in reinforcement.

Hirst, DiGennaro Reed, & Reed (2012) examined this hypothesis during an arbitrary match-to-sample task. Participants were randomly assigned to one of four groups, each associated with a level of feedback accuracy: 25%, 50%, 75%, or 100%. Inaccurate feedback—defined as receiving the feedback “Correct” following an incorrect response or “Incorrect” following a correct response—was delivered during a proportion of trials in the first phase, dependent on the group to which participants were assigned. During a second phase, only accurate feedback was given. The results showed that participants who were only provided with accurate feedback acquired the task more quickly than those who were exposed to inaccurate feedback. The percentage of correct responses during the first phase generally corresponded with the degree of feedback accuracy with higher accuracy associated with more correct responses. This negative influence of inaccurate feedback persisted even when feedback accuracy improved, suggesting that a history of inaccurate feedback influenced future responding. One hypothesis is that the inaccurate feedback resulted in the development of inaccurate rules about how to respond to the task. In the absence of contact with a discrepant contingency of reinforcement, any faulty rules that may have developed persisted. Although the authors conceptualized the study differently as an examination of integrity errors during discrete-trial instruction, DiGennaro Reed et al. (2011) used a similar procedure and had similar results. That is, when feedback and reinforcement were provided inaccurately, acquisition was negatively influenced.

Purpose and Rationale

That feedback may serve as a CSS for future responses has implications for the study of several complex human behaviors of social relevance, including the acquisition of academic responses. Although there is little research explicitly examining feedback accuracy and learning, there is a substantial body of basic and translational literature on feedback, instructional control, and rule-governed behavior. The goal of the present study is to achieve a synthesis of these literatures in an analysis of the short-and long-term effects of varying feedback accuracy on academic response acquisition in a match-to-sample task. The present study represents a systematic replication and extension of Hirst et al. (2012). The first study adopts a translational, human operant paradigm to examine the effects of feedback accuracy under highly controlled laboratory conditions using an arbitrary task. Precedence for this method of studying factors influencing learning has been set by others (e.g., Noell, Gresham, & Gansle, 2002; St. Peter Pipkin, Vollmer, & Sloman, 2010). Study 2 extends previous research by assessing the generality of findings to a less controlled, analogue educational setting and to a more applied population.

Study 1

Method

Participants and Setting

The participants for study 1 were 64 undergraduate students recruited from introductory courses in applied behavior analysis offered at a large mid-western university. Following approval from a human subjects committee for the methods and incentives used in the study participants were recruited via in-class announcements. The script used during the recruitment announcements is provided in Appendix A. Individuals responding to the announcements were predominantly female (13 males and 51 females). Participants' ages ranged from 18 to 53 ($M =$

20.06). No disabilities were reported during a demographic survey with the exception of two participants who reported an attention deficit disorder. Participants were offered extra course credit in exchange for participation amounting to 0.5% of their final grade. An exclusionary criterion was applied *post hoc* following the completion of the study. Data were excluded from analysis if the average response latency for a participant was less than 0.5 seconds during the first condition. This criterion was established to ensure participants were attending sufficiently to the task.

The study took place in a large computer lab measuring 9m by 6m. The lab contained 20 Dell OptiPlex computers each outfitted with a 17-inch wide aspect flat-panel monitor and standard mouse and keyboard. Each computer was located on a desk, arranged along the walls of the room with the monitors facing toward the center.

Materials

The stimuli constructed for the present study consisted of five nonsense shapes and nine nonsense names. The shapes were based on Japanese hiragana characters, but were significantly modified to prevent interference with future learning. Each shape was assigned a nonsense name with an additional four nonsense names created to serve as distractors. The shapes and their corresponding names are provided in Table 1.

A computer program—written in Microsoft Visual Basic.Net—was developed explicitly for the study. The program was designed to implement the entirety of the procedure. The initial interface of the program consisted of a brief demographic survey, which is provided in Appendix B. Participants were instructed to answer questions regarding gender, age, and ethnic background. Participants were also asked to self-report any physical or intellectual disability if they chose. All questions in the survey contained an option to respond with “no answer.”

Following the completion of the survey, participants were presented with written instructions on the computer screen. The instructions presented were:

Please turn off or silence all mobile devices and put them away for the duration of this study. On the next screen, you will see a symbol and a list of names. Click the button with the name associated with the symbol displayed. You should expect the task to take the entire session (about 40 to 50 minutes) so please choose carefully. If you understand these instructions, you may click the button below to begin. If you have any questions, please raise your hand and the experimenter will be available to assist you.

In order to indicate that they understood the instructions and had no questions prior to beginning the task, participants clicked a button below the instructions. This button initiated the main task and loaded the primary interface.

The program was designed to occupy the entire screen, making the other features of the computer unavailable during the study. The computer program presented an arbitrary match-to-sample task in which participants clicked a button bearing a nonsense name in response to the presentation of a nonsense shape. Shapes were presented in black within a white square measuring 6.5 cm by 6.5 cm. Five nonsense names, each displayed on a button, were presented 2.0 cm to the right of the shape in a vertical column. For reference, screenshots of the stimuli presented by the program are presented in Figures 1 and 2. For each trial—consisting of the presentation of one shape and five names followed by a response from the participant—the array of nonsense names consisted of three target names (i.e., nonsense names assigned to one of the five shapes) and two distractor names. For each trial, the name assigned to the shape being displayed was always present. The other two target names and the two distractor names were

selected at random using a function built into the programming language. The order of the names in the column was also randomized for each trial using the same method. The order in which shapes were presented was determined *a priori* using a pseudo-random sequence generated by a random number generator located at <http://www.random.org>. The sequence was pseudo-random because constraints were placed on the sequence to ensure that each shape was presented an equal number of times in each phase during the study and that shapes were never presented more than twice consecutively. The program was designed to end after the completion of 500 trials or an elapsed time of one hour.

Dependent Variables and Response Measurement

Two dependent variables were measured during the study: response to the task and trials to criterion. The participants' responses (i.e., the nonsense name selected in response to the presentation of a nonsense shape) were recorded for each trial. Responses were also coded dichotomously as either correct or incorrect. A correct response was defined as clicking the nonsense name assigned to the nonsense shape that was presented. An incorrect response was defined as clicking a nonsense name not associated with the shape, either another target name or a distractor name. Correct responses were recorded as a cumulative total of correct responses at each trial and this total was displayed graphically as a cumulative record of correct responding. The second dependent variable, trials to criterion, was defined as the number of trials completed to meet the mastery criterion, which was set at 15 consecutive correct responses.

The computer recorded all data automatically. In addition to the two dependent variables, the program also collected the latency of responses in milliseconds from the presentation of the stimuli for each trial for the purpose of applying the second exclusionary criterion.

Experimental Design and Procedure

Sixteen concurrent multiple-baseline designs across participants were conducted, each containing four participants. The multiple-baseline designs were used to compare two conditions: (a) inaccurate feedback and (b) accurate feedback. Participants were assigned to one of four groups: 25% accuracy, 50% accuracy, 75% accuracy, or a 100% accuracy comparison group, each consisting of 16 participants. The group to which a participant was assigned dictated the level feedback accuracy to which he/she was exposed during the error condition.

Upon arriving, participants were greeted and assigned randomly to a computer on which the program was already running. Following a brief introduction—presented in Appendix C—participants were instructed to begin working, starting with the demographic survey.

Inaccurate Feedback Condition. The purpose of this condition was to assess the effects of inaccurate feedback on learning. During the inaccurate feedback condition (IF), the program delivered inappropriate feedback following a proportion of responses. Two types of feedback inaccuracies were possible, depending on the participant's response. If the participant emitted a correct response, the program gave the feedback "Incorrect." If the participant emitted an incorrect response, the program gave the feedback "Correct." Both forms of feedback were displayed below the shape for 2 s followed by the presentation of the next trial. "Correct" was printed in a bold, black font printed on a green rectangle measuring 6.5 cm by 1.5 cm. "Incorrect" was printed in a bold, black font on a red rectangle of the same dimensions. Screenshots displaying both forms of feedback are presented in Figures 1 and 2.

Depending on the group to which a participant was assigned, the proportion of trials during which the program delivered inaccurate feedback varied. Participants assigned to the 25% accuracy group were exposed to inaccurate feedback during 75% of trials in IF. That is, during

an average of three out of four trials, inaccurate feedback was delivered. The feedback delivered during one out of four trials was accurate (i.e., correct responses were followed by the feedback “Correct” and incorrect responses were followed by the feedback “Incorrect”). Participants assigned to the 50% accuracy group were exposed to feedback errors during half of the trials in the error condition. Participants assigned to the 75% accuracy group were exposed to feedback errors during an average of one out of four trials in the error condition. The sequence of trials in which a feedback error was committed was predetermined as a pseudo-random sequence generated by a random number generator located at <http://www.random.org>. The randomness of the sequence was constrained by parameters ensuring that the number of trials during which inaccurate feedback was delivered was equal across all shapes.

Accurate Feedback Condition. The purpose of this condition was to assess whether prior exposure to inaccurate feedback influenced learning when the accuracy of feedback improved. During this condition, feedback errors were no longer committed. That is, all feedback delivered by the program following a response was accurate (“Correct” after a correct response and “Incorrect” after an incorrect response). No change in the program interface occurred concurrently with the change in condition. Participants continued the task until a total of 500 trials had been completed between both the IF and accurate feedback (AF) conditions. In addition, the participants assigned to the 100% accuracy comparison group were never exposed to inaccurate feedback and completed 500 trials in the AF condition.

Procedural Fidelity

Because data collection occurred automatically, the procedure was not judged to be at risk for errors in reliability or procedural fidelity (Peterson et al., 1982). However, to ensure that the program was operating optimally, algorithms were included in the program to calculate the

obtained level of feedback accuracy to ensure that participants were exposed to the conditions to which they were assigned. The obtained levels of feedback accuracy matched those programmed for all participants and conditions.

Data Analysis

Two modes of data analysis were used. Within-group analyses were conducted via visual inspection of the multiple-baseline design graphs. The graphs were cumulative records of correct responding so changes in the rate of acquisition of the task would be indicated by a change in the slope of the graph. Decisions on whether an effect was obtained were based on visual inspection of the cumulative records. That is, a difference between IF and AF was concluded if the rate of acquisition (the slope of the cumulative record) changed when and only when AF was implemented. Based upon previous data and the hypothesis that the effects of exposure to inaccurate feedback carry over into accurate feedback conditions, a decision that a carry-over effect was observed was made if the slope of the graph changes only after a delay. The effect will be replicated across participants if a delay of similar duration is observed each time.

In addition, an analysis was conducted to identify response patterns associated with the failure to acquire the tasks during IF, specifically whether inaccurate feedback was associated with acquisition of incorrect discrimination responses. An index of discrimination response strength was created by counting the frequency of each possible response (5 target names and 4 distracters) in the presence of each sample stimulus. The index of discrimination response strength was calculated by dividing these frequencies by the total number of presentations of the sample stimulus during IF. An index of 0.2 approximates chance levels of responding, but values substantially higher than 0.2 may represent the acquisition of a specific discrimination. For example, if a participant selects the response “bifdo” in the presence of the sample stimulus

associated with “zitaaf” on a high percent of trials in IF, it might be inferred that the participant has acquired an incorrect discrimination response.

Several between-group analyses were also conducted to compare the effects of the four levels of feedback accuracy. First, a percentage of correct responses was calculated for each participant in each condition to represent acquisition. Additionally, an index was created for the purposes of standardizing rates of acquisition for comparison across participants. The area under the curve (AUC) of cumulative records of correct responses was calculated using the trapezoidal method.

$$AUC = \sum (x_2 - x_1) \left[\frac{y_1 + y_2}{2} \right]$$

This value was calculated for each participant and each phase. Since the duration of each phase varied across participants as a function of the trial at which the phase change occurred, this value was standardized for comparison by also calculating the AUC of a line representing perfect responding in each phase and dividing the AUC of the cumulative record by the AUC for perfect responding, yielding a percentage value (%AUC). AUC was used in addition to the percentage of correct responses as a possibly more accurate representation of acquisition.

Two nonparametric statistical analyses were conducted to determine whether the rates of acquisition under AF were differentially influenced by the level of feedback accuracy to which each group was exposed. First, a Kruskal-Wallis ANOVA was used to determine whether the differences in %AUC across groups were due to chance. Third, a Dunn’s multiple comparison test was conducted to determine which groups’ performances differed significantly from each other.

Results and Discussion

Cumulative records for participants in the 25% accuracy group are presented in Figure 3. In addition to cumulative records of correct responses, a gray region superimposed on each graph represents the range of acquisition rates for participants in the 100% accuracy group for comparison. Acquisition was generally low, indicated by the shallow slope of the cumulative records to the left of the phase line. There was very little overlap with the data from participants in the comparison group during the initial learning curve. None of the participants met the mastery criterion during IF. After accuracy improved, acquisition of the task varied across participants. Some participants acquired the task rapidly following the removal of errors, indicated by an inflection point in the cumulative record shortly after the phase line (e.g., 102, 104, and 202). Other participants acquired the task after a longer exposure to improved instruction indicated by an inflection point further to the right of the phase line (e.g., 304, 404). Two of the 16 participants in this group did not acquire the task prior to the session ending at 500 trials (201 & 204). Of the 14 participants who did master the task, the average trial at which criterion was met was 136.43 (range: 48-220). An inspection of the cumulative records yields the conclusion that a difference was obtained between IF and AF. Specifically, the inflection point in the cumulative record occurs when and only when AF is in place. Additionally, a delay to acquisition was observed in AF, represented by the lag between the phase change and the inflection point. However, the number of trials between the phase change and the change in slope is inconsistent across participants suggesting that the magnitude of the delay is idiosyncratic.

Figure 4 displays data from participants assigned to the 50% accuracy condition. Similarly to participants in the 25% accuracy group, acquisition was low in the presence of inaccurate feedback. Two participants (208, 407) showed some overlap initially with the

comparison group, but prior to the end of IF, performances differentiated. The other 14 participants' performances were clearly differentiated from the comparison group. Also, acquisition varied similarly following the removal of feedback errors with some participants rapidly acquiring the task (e.g., 105, 106, 207, 305), while some participants acquired the task more slowly (e.g., 208, 406). Four out of 16 participants did not meet the mastery criterion (306, 307, 308, and 408). Of the 12 participants who did meet the criterion, the average trial at which criterion was met was 101.75 (range: 61-139). Visual inspection of these cumulative records yields similar conclusions to those for participants in the 25% group. None of the participants mastered the task in the presence of inaccurate feedback and the slopes of the graphs were consistently low to the left of the phase line. Changes in the slope occurred only after the phase line.

Data for participants assigned to the 75% accuracy group are presented in Figure 5. The data obtained from this group differs from those of the previous two groups. Acquisition in the presence of inaccurate feedback was higher as indicated by the steeper slopes to the left of the phase line. Four of 16 participants met the mastery criterion prior to the phase change (209, 309, 312, and 410). Additionally, acquisition for these participants fell within the range of the comparison group, suggesting that exposure to feedback errors during only 25% of trials in IF did not negatively influence learning. All but two participants quickly acquired the task following the removal of errors. However, two participants did not acquire the task (111, 112). Of the ten participants who acquired the task only after errors were removed, the average number of trials to criterion was 54 (range: 14-88). Because some participants acquired the task prior to the phase change and changes in slope were not observed between the two conditions, there is

less evidence for a difference between learning under 75% accurate feedback and 100% accurate feedback.

A conclusion was drawn that participants in the 25% and 50% accuracy groups did not acquire the task during IF. Acquisition occurred only after feedback accuracy improved and a delay to the improvement in acquisition was obtained suggesting that the negative influence of poor instruction on learning persisted under accurate feedback. Participants exposed to 75% accuracy showed higher rates of acquisition during IF with 4 participants mastering the task despite the errors. A limitation should be noted for the experimental design that may temper these conclusions. Although the design contained elements of a multiple baseline design across participants, phase changes were determined *a priori* and not based on the stability of the data making the decision rules based on baseline logic difficult to apply. In order to further support these conclusions, between-group analyses were conducted and compared to a comparison group.

The number of trials to criterion for all participants are plotted in Figure 6 with each data point representing a single participant. Data points falling below the horizontal line represent mastery prior to the phase change. Comparing the medians of the three groups, an inversely proportional relation with accuracy level was observed where lower levels of accuracy result in higher numbers of trials to criterion. The relation is somewhat weak given significant overlap across groups. However, a strong difference was observed between the groups exposed to inaccurate feedback and the 100% accuracy comparison group.

Figure 7 plots the percentage of correct responses during trials in the inaccurate feedback condition. The horizontal dotted line represents chance-level responding at 20% (5 possible responses per trial). Visually, no significant difference was obtained between participants in the

25% and 50% accuracy groups. Both groups show a low percentage of correct responses with the majority of participants responding at or below chance levels. More correct responses were emitted by participants in the 75% group, but the data are widely distributed. Participants in the 100% accuracy comparison group responded correctly most often, but some overlap was observed with the 75% accuracy group. Figure 8 displays the same data during AF. Visually, there is only a small difference between groups with the difference being primarily in distribution. Taken together, these data suggest that exposure to many feedback errors inhibited learning. Following the improvement of feedback accuracy, acquisition improved. It appears that for some participants in each group, a carryover effect was obtained, indicated by the persistence of incorrect responding in the presence of accurate feedback.

For further comparison between the groups, %AUC was calculated for each participant and each phase. Data for participants in the comparison group were arbitrarily split into two phases using the same method as for the experimental groups even though these participants were never exposed to errors. That is, all participants in the 100% accuracy condition were randomly assigned to a phase change at trial 160, 200, 260, or 300. Figures 9 and 10 present the resulting data from this calculation across all participants. Visual inspection of these graphs reveal a linear relation between feedback accuracy and %AUC with higher levels of accuracy associated with higher rates of acquisition in both conditions. However significant overlap in the data between groups was observed.

Figure 11 displays a visual representation of the analysis of whether participants acquired specific incorrect discrimination responses. Tables of all indices calculated are provided in Appendix D. Generally, participants emitted incorrect responses at or below chance levels, suggesting that the majority of participants did not acquire any incorrect discriminations in the

presence of inaccurate feedback. However, 14 data points in the 25% group and 4 data points from the 50% group fell above 0.5, suggesting that in the presence of many feedback errors, some incorrect discrimination responses were acquired. A pattern was obtained between indices of incorrect discrimination strength and accuracy level with less data points falling substantially above chance levels with higher levels of feedback accuracy.

To augment the visual inspection of these graphs in determining significant differences, nonparametric statistics were employed. A nonparametric ANOVA (Kruskal-Wallis) was used to determine whether the differences in acquisition between groups could be obtained by chance. The results of the test showed that the differences between groups were statistically significant in both phases (IF: $H(3) = 48.17, p < .0001$; AF: $H(3) = 45.53, p < .0001$). Because the Kruskal-Wallis only determines whether any of the groups were significantly different, a Dunn's Multiple Comparison Test was conducted to determine which of the groups differed significantly. Table 2 shows the results of this test for both IF and AF. During both conditions, significant differences were obtained between all groups except between 25% and 50%, and 75% and 100%. These data suggest that the number of errors to which a participant is exposed does appear to differentially influence learning. There appears to be a strong effect of many errors on acquisition, but there may be a ceiling effect where some less than perfect level of feedback accuracy does not result in stronger acquisition than perfectly accurate feedback. For example, for some participants, exposure to 75% accuracy did not appear to hinder acquisition of the task. However, for other participants, exposure to even a few errors appeared to have a negative effect. Given the distinct difference between participants 111 and 112, who did not acquire the task, and the other participants in the 75% group, it may be that the presence of even some errors significantly impacts learning. However, it may also be that the operational definition for attendance to task

developed for the exclusionary criterion did not capture some behavior of interest that could otherwise explain the data.

These findings are consistent with those in Hirst et al. (2012). Specifically, both studies found at least a weak linear relationship between the level of feedback accuracy and the rate of acquisition of the task. Further, both studies found that groups exposed to adjacent levels of feedback accuracy showed significant overlap in performances. The present study addressed some limitations to the previous study. First, the present study exposed participants to twice the number of trials in IF and the total duration of the study was approximately twice as long. This change did not appear to result in systematic differences in the data. Additionally, the present study also maintained tighter experimental control on the independent variable (obtained level of feedback accuracy) and extraneous variables (number of presentations for each shape). Again, no differences in the data were observed. Although the present study represents a systematic replication and improvement on past research, several limitations are worthy of note.

First, the generality of these findings and their applicability to educational settings may be limited. The present study was exploratory in nature and was conducted under highly controlled conditions with a population of convenience and used an entirely arbitrary task. Although the inaccurate feedback condition was extended in duration over the previous study, the exposure to inaccurate feedback was still somewhat brief. In addition, although this study was designed as an analogue to an educational setting, it is unlikely that teachers will deliver blatantly incorrect feedback. The results contribute to the literature regarding how the accuracy of feedback influences acquisition, but further research more closely modeling educational settings is needed to determine generality. For example, various applied populations may differ from undergraduate students in important ways. Additionally, there may have been a feature of

the arbitrary task that influenced learning in an unforeseen way. Thus, the purpose of study 2 was to replicate and extend the findings in an applied population, specifically typically-developing pre-school age children, using a non-arbitrary task.

Study 2

Method

Participants and Setting

The participants in this study were four pre-school age students (2 males, 2 females) recruited from a Montessori school located in a large mid-western suburb. Following human subject committee approval of the methods and incentives for the study, recruitment flyers and consent forms were sent home with students for review by parents or guardians (see Appendices E & F). In addition to consent provided by parents or guardians, verbal assent was obtained from each student prior to the start of each session. The assent procedure is provided in Appendix G. Sessions were conducted in a small room containing a large table and several chairs.

Materials

Materials for the present study consisted of four receptive tasks. Stimuli for each task were presented on 8.5” by 11” (21.6 cm x 27.9 cm) sheets of colored paper. Each sheet was placed in a clear plastic sheet protector and mounted on a three-ring binder. The first receptive task consisted of the identification of a country in Asia on a blank outline map printed on pink paper. To simplify the task, all elements of the map were blacked out except for the borders of five countries; one target and four distracters. A trial for this task consisted of an instruction, provided by the experimenter, to color in the target country using a dry-erase marker. The second task consisted of the identification of an aquatic invertebrate. Drawings of five insects were printed on a green sheet of paper. One served as the target and the other four were distracters.

During each trial for this task, the experimenter instructed the students to find the target insect and color it in or circle it. Third, students were instructed to identify, by tracing with a dry-erase marker, a river in Europe. The stimuli for this task consisted of an outline of the European continent and five thick black lines representing major rivers. Finally, the fourth task developed for the study required students to identify an image of a moon orbiting Jupiter. A colored picture of Jupiter was printed on a white sheet with five images of its moons oriented around it. In a trial, the experimenter instructed to find the target moon and circle it. Stimuli for each task are presented in Appendices H, I, J, and K and Table 3 contains the instruction script for each task and a description of the topography of corresponding responses.

The binders contained 24 copies of the stimuli for each task totaling 96 pages. The order of the pages in the binder, and the order of the tasks were determined by a pseudo-random sequence such that tasks were not presented twice consecutively and each task was presented an equal number of times. The order was randomized using a random number generator found at <http://www.random.org> and two binders were created with different sequences.

Dependent Variables and Response Measurement

The dependent variable for the present study was the participants' responses to the task. Responses were coded as either correct or incorrect. Correct responses were defined as identifying (i.e., coloring, tracing, or circling) the item corresponding with the instruction given at the beginning of the trial. Correct responses were graphed as a cumulative total. The data were graphed by blocks of 4 trials, consisting of one presentation of each task in a random order to facilitate visual inspection of the slopes of the cumulative records. The mastery criterion applied to each task was 10 consecutive correct responses. A sample data sheet is provided in Appendix L.

Interobserver Agreement and Procedural Fidelity

Interobserver agreement and procedural fidelity measures were collected for at least 30% of trials for each condition and for each participant. Interobserver agreement was collected by an independent observer and agreement was calculated for participant response as the number of agreements divided by the number of agreements plus disagreements, multiplied by 100. Interobserver agreement averaged 99.67% (range: 96.7% – 100%). Interobserver agreement ratings for each participant by condition are provided in Table 4. To measure procedural fidelity, a task analysis was created for the procedure consisting of four items the experimenter implemented during each trial. The steps included presenting the correct stimuli and the correct discriminative stimulus, providing feedback according to the programmed schedule, and presenting a token when appropriate. Procedural fidelity was measured as the percentage of steps performed by the experimenter during a session out of the total number of steps applicable to the session. Procedural fidelity averaged 99.75% (range: 98.4% – 100%). Procedural fidelity measures for each participant by condition are provided in Table 5.

Experimental Design and Procedure

The effects of four levels of feedback accuracy on learning were assessed using a multielement design embedded within an ABC design. Each of the four receptive tasks was associated with one of four levels of feedback accuracy: 25%, 50%, 75%, or 100%. The task associated with each level of accuracy was counterbalanced across participants (e.g., the task associated with 25% accuracy was different for each participant). Sessions were conducted once per day and 2 to 3 days per week. Verbal assent was obtained prior to each session and if a participant refused to assent for three consecutive sessions, he/she was excluded from the study. No participants needed to be excluded due to non-assent. A session was defined as a brief

introduction, followed by approximately 25 to 50 trials (6 to 12 trials for each task) presented in an interspersed format, and an opportunity for participants to trade in tokens earned during the session for an item from a “store,” which was populated based on the results of a preference assessment. Sessions lasted between 15 and 20 minutes.

Preference Assessment and Token Economy. A multiple stimulus without replacement preference assessment (DeLeon & Iwata, 1996) was conducted individually with each participant. Ten leisure items were included in the assessment, which were selected for academic or age appropriateness. The participant was allowed to interact with each item for 30 s prior to the beginning of the assessment. During the assessment, the 10 items were placed in a horizontal array on a table approximately 3 cm apart. The participant was seated at the table with the array approximately 5 cm in front of his or her person. The experimenter delivered the instruction, “Choose one” at the start of each trial. The chosen item was not replaced after being selected. The next trial began immediately with the same instruction being delivered by the experimenter. Trials continued until all items had been selected or until no response was made within 30 s of the instruction. In the latter case, the remaining items were coded as “not selected.” If a participant did not select at least five items, new items were selected to replace these items and the preference assessment procedure was repeated with the new array of items. The top five items selected by each participant are presented in Table 6.

The top five preferred items for each participant were used to populate a store. During a session, participants could earn tokens in the form of small colored beads, which could be exchanged for items from the store at the end of the session. A maximum of 25 tokens could be earned during a session. The prices of the items in the store were set based upon the ranking of the items in the preference assessment. The highest preferred item was worth 25 tokens and other

item prices were set such that each item was worth five less tokens than the item ranked one spot higher during the assessment. In order to track token earning, a token board consisting of an outline of a thermometer with 5 equally sized segments was placed on the table. As tokens were earned, beads were placed in one of the boxes on the thermometer. When five beads were placed in a box, the participant was permitted to color in the box. The beads were then removed and were placed in the next box and this repeated until the session ended.

Baseline. The purpose of the baseline condition was to assess whether the participants had already acquired the responses to the tasks. Each trial in baseline consisted of four components. First, the experimenter opened the binder to the first task stimulus. Next, the experimenter delivered an instruction to identify the target item for that task (e.g., “Please trace the Danube River with your marker”). Feedback was not delivered following the responses in baseline and errors were not corrected. However, participants received one token on a variable ratio 3 schedule for participation (i.e., making any response to the task and cooperating with instructions). During baseline, the token delivery was paired with a praise statement not associated with the task (e.g., “I like how you are sitting so nicely.”; “Thanks for cooperating and being a good student!”). When participants inquired about the correct answer, the experimenter informed participants “I can’t tell you yet. We have to see if you already know it first.” Participants remained in baseline until data were stable with performance at or below chance levels, or showed a consistent pattern of correct and incorrect responses for three or more consecutive presentations of each task.

Inaccurate Feedback. The purpose of the inaccurate feedback condition (IF) was to assess the effects of feedback accuracy on learning. The procedure during IF was identical to baseline except feedback was delivered following all responses to the tasks. The form of the

feedback varied depending on the level of accuracy associated with the task. During trials for the task associated with 100% accuracy, all correct responses were followed by a praise statement and one token. Incorrect responses were followed by neutral feedback (e.g., “Nice try.”) and no tokens. During trials for the task associated with 75% accuracy, feedback was delivered identically to the 100% accuracy task except that on one of four trials, inaccurate feedback was given. Inaccurate feedback was defined as following a correct response with a neutral statement and no tokens or following an incorrect response with a praise statement and one token. During trials for the 25% accuracy task, the procedure was reversed such that three of four responses were followed by inaccurate feedback. Finally, during trials for the 50% accuracy task, the type of feedback was scheduled such that half of responses were followed by inaccurate feedback. Errors were not corrected in this condition regardless of accuracy level.

Accurate Feedback. The purpose of this condition was to assess whether prior exposure to inaccurate feedback influenced learning even after accuracy improved. The procedure for the accurate feedback (AF) condition was identical to the procedure for IF except that feedback errors were no longer committed. That is, all correct responses were followed by a praise statement and one token while all incorrect responses were followed by a neutral statement and no tokens.

Data Analysis

Data on correct responses for each task were graphed as the cumulative total of correct responses. A slope of 1.0 represents perfect acquisition of the task while slopes lower than 1.0 represent lower rates of acquisition. The decision on whether an effect was obtained was made by applying baseline logic during visual inspection of the multi-element design graphs. For the present study, the decision on whether an effect was obtained will be made if the data series for

each level of treatment integrity are consistently differentiated over time. If the data series are consistently differentiated, the conclusion will be drawn that the conditions influence behavior differently. However, if the series are not differentiated, it will not be concluded that the conditions influenced behavior differently.

An analysis was also conducted to determine whether the failure to acquire tasks during IF was associated with specific patterns of responses. The frequency of each possible response (1 target, 4 distractors, and responses other than the five programmed comparison stimuli) was counted for each task. An index of discrimination strength was calculated by dividing the frequency of each selection by the number of presentations of the task during IF. Since there were 5 programmed comparison stimuli for each task, an index of 0.20 approximates chance levels. Indices substantially above 0.20 for incorrect responses may represent the acquisition of an incorrect discrimination response.

Results and Discussion

Figure 12 depicts data for participant A. During baseline, correct responses were emitted near or below chance levels, indicating the tasks had not already been mastered. During IF, the task associated with 100% accurate feedback was mastered at block 19, and the tasks associated with inaccurate feedback continued with a low proportion of correct responses. No observable differentiation between the inaccurate feedback conditions was observed. During AF, acquisition of the tasks previously associated with inaccurate feedback occurred. The first to meet mastery criterion was the task previously associated with 50% accuracy after 22 trials at block 50. The 25% accurate feedback task was mastered after 3 additional trials at block 53. Last, the task previously associated with 75% accuracy was mastered after 26 trials at block 54. After a period of 4 weeks, a maintenance probe was conducted. For all tasks except the task previously

associated with 25% accuracy, correct responses were emitted for all blocks. For the 25% accuracy task, correct responses were emitted during the first 3 blocks, followed by three incorrect responses during the last 3 blocks.

Data for participant B are depicted in Figure 13. There were no correct responses emitted during baseline, indicating that the tasks had not been previously learned. During IF, participant B did not demonstrate acquisition of any tasks initially. Correct responses were emitted near or below chance levels for approximately 144 blocks. At block 119, the neutral statement following an incorrect response to the 100% accuracy tasks was modified to “No, that’s not it. Maybe try picking a different one next time.” This change to the procedure was made because participant B was consistently emitting the same incorrect response and not sampling the contingencies during the 100% accuracy task. The feedback delivered in its original format did not appear to influence her behavior. That is, receiving feedback that her response was incorrect did not result in varied responding on future trials. This pattern was not observed for the other three tasks. The block at which the feedback was changed is depicted by an asterisk on the graph. At block 159, the mastery criterion for the 100% accuracy task was met. No other tasks were mastered during IF. During AF, acquisition of the tasks previously associated with inaccurate feedback was observed. The 75% accuracy task showed acquisition first, but incorrect responses were still emitted until the mastery criterion was met after 63 trials at block 229. The task associated with 50% accuracy was mastered after 66 trials at block 232. The 25% accuracy task was mastered after 68 trials at block 234. During a maintenance probe, correct responding continued for all tasks.

Figure 14 portrays data for participant C. Correct responding occurred at or below chance levels during baseline. During IF, the task associated with 100% accuracy was mastered in 11

trials, meeting criterion at block 19. The tasks associated with inaccurate feedback were not acquired during IF and little differentiation between the tasks was observed. During AF, the first task to be mastered was the task previously associated with 75% accuracy, meeting criterion in 10 trials at block 33. The 25% accuracy task was mastered in 19 trials at block 42. The task previously associated with 50% accuracy was mastered in 34 trials at block 57. During a maintenance probe, correct responses were emitted for all tasks except the task previously associated with 50% accuracy. No correct responses were emitted for this task.

Data for participant D are displayed in Figure 15. During baseline, no correct responses were emitted. During IF, the 100% accuracy task was mastered in 14 trials at block 18. The tasks associated with inaccurate feedback were not acquired with only one correct response emitted between the three tasks. During AF, the first task to be mastered was the task previously associated with 75% accuracy in 11 trials at block 31. Second, the task previously associated with 25% accuracy was mastered in 14 trials at block 34. The 50% accuracy task was mastered last after 23 trials at block 43. During the maintenance probe, participant D emitted correct responses for the task previously associated with 50% accuracy. Three correct responses out of four blocks were emitted for the 100% accuracy task. Out of four blocks, two correct responses were emitted for the 25% accuracy task and one for the 75% accuracy task.

The results demonstrate a clear difference between learning under 100% accurate feedback and any amount of inaccurate feedback. That is, task acquisition occurred the most rapidly and consistently when only accurate feedback was provided. However, a clear relation was not obtained for the three levels of inaccurate feedback. During IF, acquisition was largely undifferentiated between the three imperfect feedback conditions for three of four participants. At the end of IF, participant B showed a differentiation in the expected pattern, with the most

correct responses emitted to the 100% accurate feedback task and the least to the 25% accurate feedback task, although the difference between them was minimal. In contrast, the other participants showed no differentiation or even a slight differentiation counter to expectations (participant C). The relation between the level of feedback accuracy and the magnitude of delay obtained in AF before the task was mastered was also not clear. Again, participant B acquired the three inaccurate feedback tasks in the expected order with the highest level of accuracy associated with the shortest delay to mastery (75% accuracy). However, this pattern did not occur with the other three participants. Finally, during maintenance probes, the level of feedback accuracy previously associated with the tasks did not appear to systematically influence maintenance. Participant A failed to maintain mastery of the task previously associated with 25% accurate feedback. However, participant D showed the lowest maintenance of the task associated with 75% accurate feedback and for participant C, the 50% accurate feedback task.

A consistent pattern was obtained, however, on the basis of the task. That is, when the river task was associated with any level of inaccurate feedback, maintenance of the task was low. In addition, when the river task was associated with 100% accurate feedback (participant B) incorrect responses were still occasionally emitted even after the task had met mastery criterion. It may be that the river task was qualitatively more difficult than the other three tasks, that is, the relevant features of the stimuli may have been less discriminable.

The results of the analysis of indices of incorrect discrimination response strength are depicted graphically in Figure 16. The top panel depicts the strength of incorrect discrimination responses as represented by the proportion of selections during IF. Generally, it appears that participants acquired incorrect discrimination responses to the tasks associated with inaccurate feedback. That is, participants selected the same incorrect response for which they occasionally

received praise and a token until AF was implemented. The analysis showed that this did not occur during the task associated with only accurate feedback. The bottom panel depicts the same data, but organized by task. This analysis was used to determine whether incorrect discriminations learned were better predicted on the basis of condition or task. The data support the conclusion that a consistent pattern for the acquisition of incorrect discriminations was obtained by condition. Variability was approximately equal across tasks.

The results of the present study are consistent with previous research and with the results of study 1 in that the best learning outcomes were obtained when only accurate feedback was delivered. The lack of differentiation during IF between the three levels of inaccurate feedback appears to be consistent with the overlap between groups in study 1. DiGennaro Reed et al. (2011) also showed little differentiation between conditions in which various levels of errors were committed. Although the results of the present study extend the line of research on feedback accuracy, a few limitations should be noted. Although the present study was designed to more closely approximate an educational setting, it is unlikely that teachers will consistently deliver blatantly inaccurate feedback. Additionally, the present study evaluated the acquisition of a non-arbitrary task but is not likely representative of the range of academic tasks to which students are exposed.

General Discussion

The purpose of study 1 was to evaluate the effects of feedback accuracy on learning under highly-controlled laboratory conditions, and to systematically replicate and extend the findings of previous research. Study 2 was an extension of this line of research to a setting and population more closely mimicking an applied educational context to assess the generality of the effects outside of the laboratory. The present studies yielded similar results. When the feedback

delivered following responses was inaccurate, participants failed to acquire the tasks. Both data sets demonstrate at least a weak linear relation between feedback accuracy and acquisition rate. The highest rates of acquisition occurred during 100% feedback accuracy and the lowest rates occurred during 25% feedback accuracy. It is important to note, however, that the differences between the three conditions associated with inaccurate feedback do not appear to represent clinically meaningful differences and a significant amount of overlap across conditions was observed in both studies.

A similar finding was also obtained for the analysis involving indices of discrimination response strength. In both studies, the delivery of feedback at 25% and 50% accuracy appears to support the acquisition of some incorrect discriminations as represented by participants selecting a specific incorrect response consistently in the presence of a sample stimulus. For example, participant 202 (study 1, 25% accuracy group) selected “bifdo” in response to 90% of the presentations of the shape associated with “zitaaf.” Similarly, participant A selected the Dnieper River in response to 63.6% of the presentations of the discriminative stimulus “Danube River”, which was the task associated with 25% feedback accuracy. Many other incorrect discrimination responses were emitted at or below chance levels in both studies, suggesting that inaccurate feedback does not consistently support the acquisition of incorrect responses in all participants. For participants who did not appear to acquire an incorrect discrimination response, response patterns were more consistent with random patterns of responding.

Comparisons between data from both studies during AF also reveal similarities. Measures of acquisition for the tasks previously associated with 25% and 50% accuracy were not differentiated in either study. The data were widely distributed across participants and a comparison of distribution via ranges of percent correct responses emitted during AF revealed

considerable overlap across studies. Additionally, the carryover effect of prior exposure to inaccurate feedback was replicated across studies. That is, a carryover effect was obtained for some, but not all, participants across conditions and studies illustrated by a delay to acquisition following the removal of feedback errors.

Due to the consistency of findings across studies, several conclusions have been supported. First, the most desirable learning outcomes were obtained when students were never exposed to inaccurate feedback. Initial acquisition occurred most rapidly under 100% accurate feedback and acquisition was consistently maintained both across sessions and during maintenance probes. For some undergraduate student participants in study 1, slight deviation from the prescribed feedback procedure as represented by the 75% accuracy condition did not result in a significantly reduced rate of acquisition. However, some participants showed a high sensitivity to the delivery of occasionally inaccurate feedback, which was especially apparent in the data from child participants in study 2. Third, exposure to many trials of inaccurate feedback, represented by the 25% and 50% accuracy conditions, consistently resulted in failures to acquire correct discrimination responses. Fourth, for participants who did not master the discrimination tasks, responses to the tasks were consistent with one of two patterns. Participants either responded randomly suggesting that exposure to inaccurate feedback resulted in a failure to acquire the correct response, or participants responded in a pattern consistent with the acquisition of a specific incorrect discrimination. Finally, with the exception of the conclusion that learning occurred most reliably when feedback provided was accurate, the results of both studies show a high degree of idiosyncrasy in results across participants.

Furthermore, the results are consistent with previous research on feedback accuracy, specifically those of Hirst et al. (2012). The weak linear relation between accuracy and task

acquisition was replicated in study 1 and the results were extended by providing a more in-depth analysis of response patterns associated with learning under conditions of feedback inaccuracy. Study 2 yielded similar results in a more applied context supporting the generality of the findings across settings and populations. Of note was that when child participants were exposed to inaccurate feedback, a more pronounced effect was obtained with all participants failing to acquire the task associated with 75% accurate feedback compared to the 4 participants in study 1 who met the mastery criterion during IF.

The present studies contribute to the literature on performance feedback, instruction, and staff training. Firstly, these findings provide a foundation for the examination of accuracy as a dimension of performance feedback. Meta-analyses and reviews have compiled a number of feedback characteristics and dimensions that have been shown to influence its efficacy (e.g., Balcazar, Hopkins, & Suarez, 1985). Among these characteristics are source, medium, privacy, and content (Alvero et al., 2001). The results of the present studies suggest that the accuracy of feedback might be included as well. Although the method used to evaluate feedback accuracy represents an extreme case of inaccurate feedback in the form of blatantly incorrect feedback, variations and lesser forms of feedback inaccuracy may occur in organizational settings. For example, if an individual receives feedback from two supervisors who differentially apply criteria for correct and incorrect performances, an individual may receive both positive and corrective feedback for the same behavior. Students receive feedback often in educational settings either in written form, verbal feedback, or through numeric or letter grades. When a learning task has well-defined correct and incorrect responses, it may be unlikely that students will encounter inaccurate feedback or conflicting feedback from two sources. However, it might be more probable that inaccurate or conflicting feedback could be delivered following responses

to more abstract, creative, or conceptual tasks. The present studies represent a beginning in a line of research necessary to delineate this dimension.

Second, an interpretation of the carryover effect obtained during AF might be informed by the instructional control literature. The results of several studies have shown that instructional control of behavior, under some conditions, can prevent individuals from contacting underlying reinforcement contingencies, resulting in a failure to maximize reinforcement (e.g., Galizio, 1979). The function of the feedback provided in both studies was not explicitly evaluated, but it is likely that the feedback served multiple functions. One function that has been suggested for feedback is a contingency specifying stimulus (Agnew & Redmon, 1993). That is, participants may have developed a set of rules following receipt of feedback such that if feedback was received indicating the preceding response was correct, the contingency specified is that in the same context in the future, the same response will be reinforced. This assumes that receiving positive feedback serves as a reinforcer. An opposite rule might be specified when feedback is received that the preceding response was incorrect. An interpretation based on these findings would suggest that participants developed rules during IF and continued to follow those rules during AF when the contingencies had changed. The carryover effect, or the delay before participant responses changed to correspond to the new contingencies in place may represent the rigidity of rule-governed behavior described by studies on instructional control (e.g., Shimoff, Catania, & Matthews, 1981).

The present studies were designed as analogues for educational settings. Implications for educational practices and teacher training may be drawn. As these results and the results of parametric analyses of treatment integrity or instructional fidelity have demonstrated, the most desirable educational outcomes are typically achieved when procedures are implemented

precisely as intended. Although a growing body of literature has brought the necessity of fidelity monitoring to light in the area of treatments for undesirable behaviors (e.g., Northup et al., 1997; Vollmer et al., 1999; St. Peter Pipkin et al., 2010), fewer studies have focused on the fidelity of instructional procedures, although this area is growing as well (e.g., Noell, Gresham, & Gansle, 2002; Holcombe, Wolery, & Snyder, 1994; Grow et al., 2009; DiGennaro Reed et al., 2011; Hirst et al., 2012). These data support the importance of ensuring high levels of implementation during instruction in order to achieve the best learning outcomes. There are implications for teacher and staff training, specifically for mastery criteria applied during initial and follow-up training. The staff training literature has several recommendations for designing effective training programs. One such recommendation is that training be performance- and criterion-based, which is defined as a continuing training effort ending only when staff demonstrate the criterion level of performance targeted (Reid & Parsons, 2002). The literature contains a wide variety of mastery criteria for implementation proficiency on the part of staff or teachers. In the context of the implementation of behavior support plans, Reid and Parsons (2002) have suggested the *80% Rule*, that is, a behavior support plan implemented at 80% is likely to be effective (p. 204). The authors refer to this rule as more of a guideline as individual support plans may require a higher level of implementation and criterion performance should be determined on a case-by-case basis. Parametric analyses of fidelity and feedback procedures may serve to inform guidelines for staff training such as the *80% Rule*. The results of the present study and other aforementioned studies suggest that in a small number of cases, 75% implementation may result in desired outcomes with adult participants, but the data do not support the same conclusion for children. It may be the case that 80% implementation or accuracy may be sufficient in some cases. However, further titrations are necessary in both lines of research to

determine if there is a level of accuracy or fidelity that yields outcomes similar enough to perfect levels. It is not likely that these parallel lines of research will arrive at a solid criterion given the idiosyncrasy of findings across participants, until the variables that result in individual differences are identified.

The present studies also yielded results consistent with the results of parametric analyses of treatment integrity, which may have implications for future research on feedback accuracy. DiGennaro Reed et al. (2011) was framed as a parametric analysis of varying levels of treatment integrity during a discrete trial instruction procedure. Although the purpose of the present analysis was framed around feedback accuracy, similar behavioral principles or processes may be involved. Some have argued that feedback can serve a reinforcement function (e.g., Peterson, 1982). For individuals who have a long history of reinforcement for responding to feedback, the feedback might acquire this function as a conditioned reinforcer. Although the present studies did not evaluate whether the feedback provided functioned as a reinforcer, if this function were found, then the methodology of the present studies closely resembles that of several studies in the treatment integrity literature that experimentally manipulated reinforcement procedures. Deviations from prescribed feedback procedures may therefore be included among the numerous errors already examined in the integrity literature.

A relation between feedback accuracy and treatment integrity, if confirmed, has implications for future research. For example, past parametric analyses of integrity have shown sequence effects, with a reduced negative impact of degradations in treatment integrity if integrity was initially high. St. Peter Pipkin et al. (2010) showed that if the integrity of a procedure designed to reduce problem behavior was reduced following a period of perfect implementation, a lesser impact on behavior was obtained. Northup et al. (1997) showed that

when a treatment was initially implemented perfectly, treatment effects were maintained until implementation strength reached 20%. Future research might examine whether similar sequence effects are obtained for the accuracy of feedback. This might prove to be important information for educators and staff trainers. Follow-up strategies are important components of staff training procedures as the implementation of new skills tends to decay between training and the opportunity to implement the procedure in the classroom (Noell, Duhon, Gatti, & Connell, 2002). Following initial training, staff often require follow-up coaching to shape and maintain implementation behaviors with fidelity (e.g. Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). In the present studies, when initial feedback accuracy was low, a persistent effect was obtained even after accuracy improved. Although in both studies the carryover effect was somewhat short-lived, delays accumulating as a result of consistently inaccurate feedback across learning tasks could result in a significant delay in a student's overall learning (Grow et al., 2011). If initially accurate feedback can be demonstrated to reduce the negative impact of future degradations in accuracy, the necessity of resource-intensive follow-up strategies may be reduced and negative impacts on student learning might be prevented by ensuring initial accuracy.

There are some limitations to the present studies that serve to temper these conclusions, but they also provide directions for future research. There is only a limited pool of extant research on the topic of feedback accuracy so the present studies were exploratory in nature. As such, they were designed to examine the effects of experimentally manipulating feedback under ideal conditions in a translational, human operant paradigm. Because the studies were conducted in analogue settings, the generality of these findings to applied educational settings may be limited. Future research might more closely emulate these applied settings to test for generalization. Secondly, the experimental manipulation of feedback accuracy represented an

extreme case of inaccurate feedback with feedback delivered that was dichotomously opposed to the prescribed feedback. It is unlikely that teachers would consistently deliver blatantly incorrect feedback to students at the high rates assessed in the present studies. As previously discussed, it may be more likely that feedback inconsistency would occur being delivered by two different sources. Third, although study 1 extended the exposure to inaccurate feedback to which participants were exposed compared to Hirst et al. (2012), the exposure was still relatively brief and the results of both studies remained comparable suggesting that a modest increase in exposure did not augment the effect. It may be, however, that a significant increase in the duration of exposure to inaccurate feedback may prolong the carryover effect obtained.

Some supplemental observations made during study 2 may also provide directions for future research. In Figure 6, the number of trials to criterion for participants in study 1 are plotted and a weak inversely proportional relation between accuracy level and the number of trials to criterion with higher levels of accuracy resulting in fewer trials before the mastery criterion was met. This relation was not replicated in study 2. One might predict that the tasks previously associated with higher levels of feedback accuracy would be acquired first, but this was not typically the case. For example, participant A acquired the task associated with 50% accurate feedback first out of the three inaccurate feedback tasks and the 75% accuracy task last. Additionally, participants C and D acquired the 25% accuracy task before the 50% accuracy task. Given the degree of overlap in trials to criterion across the three inaccurate feedback conditions in study 1, it is not overly surprising that the sequence of acquisition among the relatively small sample of participants in study 2 would be inconsistent. Also, conclusions about the sequence of acquisition may have been confounded by heuristics applied by participants. Participant A tacted such a rule during the session preceding the session during which he met criterion for all three

inaccurate feedback tasks. That is, after the contingencies had changed, he tacted that he would try each of the possible responses until he got them right. He then proceeded to select each possible response systematically moving from right to left on each page. Because of this systematic pattern of responding, he contacted accurate, positive feedback for the 25% and 50% accuracy tasks first because both target stimuli were on the right side of their respective arrays. The other three participants did not overtly tact any strategies during AF, but it may be the case that the acquisition sequence was confounded for similar reasons.

A second observation might direct future research. Participant D in study 2 frequently tacted a discrepancy in feedback when the two different forms of feedback possible followed the same response on consecutive presentations of a task. For example during the moon task, the discriminative stimulus “Io” was presented and participant D selected Europa (one of the distracters) and received feedback indicating that was the correct answer. On the next presentation of the moon task, she selected Europa again and received feedback that it was an incorrect answer. On several occasions where this pattern of responding and feedback occurred, the participant would ask, “Why do you keep doing that?” The other three participants did not explicitly tact the salience of the inconsistent feedback delivered during IF. It is possible that the inconsistent feedback presented resulted in a self-generated rule that may have influenced responding during the tasks. Her behavior is suggestive of a lower sensitivity to the feedback provided than other participants given the prolonged number of trials required to reach mastery criterion even for the task associated with 100% accurate feedback. An empirical question that might be examined in the future is whether and to what extent self-generated rules influence sensitivity or compliance to feedback delivered by an outside source.

Both a limitation and a direction for future research along the current line is the interaction of task difficulty and inaccurate feedback. Study 2 used a counter-balanced multielement design to examine the effects of four levels of inaccurate feedback on acquisition within-subjects. Because four different tasks were used, it is possible that some of the effects may have been influenced by differential difficulty between the four tasks, although an effort was made to control for this variable. A difference in maintenance of learning was observed across participants in study 2. This difference was better predicted on the basis of task (the river task) than by accuracy level, which suggests that the task itself may have influenced the rate of learning. It may be the case that less accurate feedback is required to promote acquisition for easy tasks than for more complex or difficult tasks.

Finally, feedback delivered in organizational and educational settings is not always delivered alone. Reviews of the characteristics of feedback that influence efficacy have shown that feedback combined with other procedures such as reinforcement results in more consistent positive effects than feedback alone (Alvero et al., 2001). Although, study 2 incorporated a putative conditioned reinforcement procedure by including a token economy, which was manipulated along with feedback errors, the data obtained are insufficient to determine whether the tokens actually served as reinforcers during the study. Future research might examine whether the addition of consequent procedures, such as positive or negative reinforcement, to a manipulation of feedback accuracy influences acquisition.

Conclusion

The present studies evaluated the short- and long-term effects of varying levels of feedback accuracy on the acquisition of a match-to-sample task. The results support the conclusion that exposure to inaccurate feedback results in a delay to acquisition, which persists

after the accuracy of feedback improves. The results were consistent across experimental preparations and replicated the effects found in previous research. Although little research has examined feedback accuracy explicitly, the similarities between the present studies and the extant literature on treatment integrity and instructional control provide a basis for an interpretation of the results from several perspectives. Further research is needed to determine whether these similarities are representative of similar behavioral processes. These results and those of future research may help to inform understanding of the function of feedback and its dimensions, which may have implications for the use of feedback in educational or organizational settings.

References

- Agnew, J. L., & Redmon, W. K. (1993). Contingency specifying stimuli: The role of “rules” in organizational behavior management. *Journal of Organizational Behavior Management*, *12*, 97-76. doi: 10.1300/J075v12n02_04
- Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management*, *21*, 3-29. doi: 10.1300/J075v21n01_02
- American Psychological Association (n.d.). *Ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/code/principles.pdf>
- American Psychological Association, Division of Clinical Psychology, Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, *48*, 3-23.
- Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, *7*, 65-89.
doi: 10.1300/J075v07n03_05
- Behavior Analyst Certification Board (n.d.) *Guidelines for responsible conduct for behavior analysts*. Retrieved from <http://www.bacb.com/Downloadfiles/BACBguidelines/1007GuidelinesFpdf.pdf>
- Billingsley, F., White, O. R., & Munson, R. (1980). Procedural reliability: A rationale and example. *Behavioral Assessment*, *2*, 229-241.
- Boruch, R. F., & Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluations. *Professional Psychology*, *8*, 411-434. doi: 10.1037/0735-7028.8.4.411

- Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing & Health, 33*, 164-173. doi: 10.1002/nur.20373
- Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., ..., Woody, S. R. (1998). Update on empirically validated treatments II. *The Clinical Psychologist, 51*, 3-16.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7-18. doi:10.1037/0022-006X.66.1.7
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: controversies and evidence. *Annual Review of Psychology, 52*, 685-716. doi:10.1146/annurev.psych.52.1.685
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45. doi:10.1016/S0272-7358(97)00043-3
- Daniels, A. C. (1994). *Bringing out the best in people*. New York: McGraw-Hill.
- DeGrandpre, R. J., & Buskist, W. F. (1991). Effects of accuracy of instructions on human behavior: Correspondence with reinforcement contingencies matters. *The Psychological Record, 41*, 371-384.
- Detrich, R. (2008). Evidence-based, empirically supported, or best practice? A guide for the scientist-practitioner. In J. K. Luiselli, D. C. Russo, W. P. Christian, & S. Wilczynski (Eds.): *Effective practices for children with autism: Educational and behavior support interventions that work*. New York: Oxford Press.

- Dib, N., & Sturmey, P. (2007). Reducing student stereotypy by improving teachers' implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis, 40*, 339-343. doi:10.1901/jaba.2007.52-06
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40*, 447-461. doi:10.1901/jaba.2007.40-447
- DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review, 34*, 220-231.
- DiGennaro Reed, F. D., Miller, J. R., Hirst, J. M., Reed, D. R., & Kaplan, B. A. (2012). Choose your words wisely: phrasing influences adherence to unprofitable rules. Manuscript submitted for publication.
- DiGennaro Reed, F. D., & Reed, D. R. (2008). Towards an understanding of evidence-based practice. *Journal of Early and Intensive Behavior Intervention, 5*, 20-29.
- DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis, 44*, 611-615. doi:10.1901/jaba.2011.44-611
- Duhon, G. J., Mesmer, E. M., Gregerson, L., & Witt, J. C. (2009). Effects of public feedback during RTI team meetings on teacher implementation integrity and student academic performance. *Journal of School Psychology, 47*, 19-37. doi:10.1016/j.jsp.2008.09.002

- Duncan, P. K., & Bruwelheide, L. R. (1986). Feedback: Use and possible behavioral functions. *Journal of Organizational Behavior Management*, 7, 91-114.
doi:10.1300/J075v07n03_06
- Eckert, T. L., Dunn, E. K., & Ardoin, S. P. (2006). The effects of alternate forms of performance feedback on elementary-aged students' oral reading fluency. *Journal of Behavioral Education*, 15, 149-162. doi:10.1007/s10864-006-9018-6
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005).
Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Ford, J. E., (1980). A classification system for feedback procedures. *Journal of Organizational Behavior Management*, 2, 183-191. doi:10.1300/J075v02n03_04
- Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior*, 31, 53-70.
doi:10.1901/jeab.1979.31-53
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011).
Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31, 79-88.
doi:10.1016/j.cpr.2010.09.007
- Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, R. (1993) Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, 22, 254-272.

- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*, 198-205. doi:10.1207/SLDRP1504_4
- Grow, L. L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonsalves, L., Ktaech, I. A., & Kisamore, A. N. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity. *Journal of Behavioral Education, 18*, 142-156. doi:10.1007/s10864-009-9085-6
- Hirst, J. M., DiGennaro Reed, F. D., & Reed, D. D. (2012). Effects of varying feedback accuracy on task acquisition: A computerized translational study. *Journal of Behavioral Education*. Advance online publication. doi: 10.1007/s10864-012-9162-0
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 683-687. doi:10.1037//0022-006X.63.5.683
- Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4*, 49-73. doi:10.1007/BF01560509
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.
- Individuals with Disabilities Education Improvement Act of 2004, H.R. 1350, 108th Congress.
- Kaufman, A., Baron, A., & Kopp, R. E. (1966). Some effects of instructions on human operant behavior. *Psychonomic Monograph Supplements, 1*, 243-250.

- Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A behavioral approach to occupational safety: Pinpointing and reinforcing safe performance in a food manufacturing plant. *Journal of Applied Psychology, 63*, 434-445. doi:10.1037//0021-9010.63.4.434
- Lane, K. L., Jolivette, K., Conroy, M., Nelson, C. M., & Benner, G. J. (2011). Future research directions for the field of E/BD: Standing on the shoulders of giants. *Education and Treatment of Children, 34*, 423-443. doi:10.1353/etc.2011.0029
- 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 behavior analysis 1991 2005. *Journal of Applied Behavior Analysis, 40*, 659-672. doi:10.1901/jaba.2007.659-672
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266. doi:10.1016/0272-7358(91)90103-2
- National Autism Center (n.d.). *National standards report*. Retrieved from <http://www.nationalautismcenter.org/nsp/reports.php>
- No Child Left Behind Act of 2001, H.R. 1, 107th Congress.
- Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). Mahwah, NJ: Erlbaum.
- Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up, and implementation of behavior management interventions in general education. *School Psychology Review, 31*, 217-234.

- Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*, 51-67. doi:10.1023/A:1014385321849
- Northup, J., Fisher, W., Kahng, S., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities, 9*, 1-16. doi:10.1023/A:1024984526008
- Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice, 18*, 148-153. doi:10.1111/j.1468-2850.2011.01246.x
- Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology, 77*, 212-218. doi:10.1037/a0015232
- Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst, 5*, 101-102.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.
doi:10.1901/jaba.1982.15-477
- Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management, 3*, 1-16.
doi:10.1300/J075v03n01_01
- Reid, D. H., & Parsons, M. B. (2002). Essential procedures for training staff to implement behavior support plans. In D. H. Reid & M. B. Parsons (Eds.), *A specific process for monitoring staff implementation of behavior support plans*. Morganton, NC: Habilitative Management Consultants, Inc.

- Ribes, E., & Rodriguez, M. E. (2001). Correspondence between instructions, performance, and self-descriptions in a conditional discrimination task: The effects of feedback and type of matching response. *The Psychological Record, 51*, 309-333.
- Ribes-Iñesta, E. (2000). Instructions, rules, and abstraction: A misconstrued relation. *Behavior and Philosophy, 28*, 41-55.
- Sackett, D. L., Rosenberg, W. M. C., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ, 312*, 71-72.
doi:10.1136/bmj.312.7023.71
- Sanetti, L. M. H., & DiGennaro Reed, F. D. (2012). Barriers to implementing treatment integrity procedures in school psychology research: Survey of treatment outcome researchers. *Assessment for Effective Intervention, 37*, 195-202. doi:10.1177/1534508411432466
- Sanetti, L. M. H., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational and Psychological Consultation, 21*, 209-232.
doi:10.1080/10474412.2011.595163
- Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*, 445-459.
- Schlinger, H., & Blakely, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior Analyst, 10*, 41-45.
- Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*, 460-475.

- Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior*, 36, 207-220.
- Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. Englewood Cliffs, NJ: Prentice-Hall
- Song, M. K., Happ, M. B., & Sandelowski, M. (2010). Development of a tool to assess fidelity to a psycho-educational intervention. *Journal of Advanced Nursing*, 66, 673-682.
doi:10.1111/j.1365-2648.2009.05216.x
- St. Peter Pipkin, C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis*, 43, 47-70. doi:10.1901/jaba.2010.43-47
- Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. M., Iwata, B. A., & Lovaas, O. I. (1988). The right to effective behavioral treatment. *Journal of Applied Behavior Analysis*, 21, 381-384. doi:10.1901/jaba.1988.21-381
- 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. doi:10.1901/jaba.1999.32-9
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620-630. doi:10.1037//0022-006X.61.4.620
- What Works Clearinghouse (n.d.). *Procedures and standards handbook (Version 2.1)*. Retrieved from <http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19>

- 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. doi:10.1901/jaba.2006.144-05
- Wolery, M. (1994). Procedural fidelity: A reminder of its functions. *Journal of Behavioral Education, 4*, 381-386. doi:10.1007/BF01539539
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156-167. doi:10.1037//0022-006X.49.2.156

Table 1

Target shapes and corresponding names

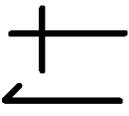
Name	Zitaaf	Raopol	Smuzy	Punfi	Bifdo
Shape					

Table 2

Significance levels of Dunn's Multiple Comparison Test. Data above the diagonal represent significance levels of differences between groups in the inaccurate feedback condition. Data below the diagonal represent significance levels between groups during the accurate feedback condition. Non-significant differences are marked as ns.

Group	25% Accuracy	50% Accuracy	75% Accuracy	100% Accuracy
25% Integrity	-	ns	.0001	.0001
50% Integrity	ns	-	.001	.0001
75% Integrity	.0001	.01	-	ns
100% Integrity	.0001	.0001	ns	-

Table 3

List of instructions delivered for each task. For instructions not specifying a specific response topography, any discrimination response was accepted as a response.

Task	Instruction	Accepted Responses
Insects	“Which one is the <target>?” “Where is the <target>?” “Circle the <target>.” “Color in the <target>.” “Find the <target>.”	Circle with marker, color with marker, touch with marker, point with finger, touch with hand
Rivers	“Trace the <target>.” “Which one is the <target>?” “Where is the <target>?” “Find the <target>.” “Circle the <target>.”	Circle with marker, color with marker, touch with marker, point with finger, touch with hand
Countries	“Color in <target>.” “Where is <target>?” “Which one is <target>?” “Circle <target>.” “Find <target>.”	Circle with marker, color with marker, touch with marker, point with finger, touch with hand
Moons	“Circle <target>.” “Which one is <target>?” “Where is <target>?” “Find <target>.”	Circle with marker, color with marker, touch with marker, point with finger, touch with hand

Table 4

Interobserver agreement by participant and condition

Participant	Condition				Total
	Baseline	Inaccurate Feedback	Accurate Feedback	Maintenance Probe	
A	100%	96.70%	100%	100%	98.59%
B	100%	99.12%	98.89%	100%	99.15%
C	100%	100%	100%	100%	100%
D	100%	100%	100%	100%	100%

Table 5

Procedural fidelity by participant and condition

Participant	Condition				Total
	Baseline	Inaccurate Feedback	Accurate Feedback	Maintenance Probe	
A	100%	99.58%	100%	100%	99.82%
B	100%	99.12%	100%	100%	99.44%
C	98.44%	100%	100%	100%	99.58%
D	100%	100%	98.89%	100%	99.55%

Table 6

Preferred items identified during the preference assessment for each participant

Participant	Items selected
A	<ol style="list-style-type: none"> 1. rubber popper 2. zoo animal sticker 3. star sticker 4. animal eraser 5. pencil grip
B	<ol style="list-style-type: none"> 1. colorful pencil 2. star-shaped eraser 3. tie-dye pencil 4. pencil grip 5. box of 4 crayons
C	<ol style="list-style-type: none"> 1. zoo animal sticker 2. star sticker 3. star eraser 4. pencil grip 5. rubber popper
D	<ol style="list-style-type: none"> 1. monster finger puppet 2. rubber popper 3. zoo animal sticker 4. box of 4 crayons 5. star sticker

Figure Captions

Figure 1. Screenshot of the computer interface during the match-to-sample task displaying “Correct” feedback in green below the sample stimulus.

Figure 2. Screenshot of the computer interface during the match-to-sample task displaying “Incorrect” feedback in red below the sample stimulus.

Figure 3. Cumulative records for participants in the 25% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.

Figure 4. Cumulative records for participants in the 50% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.

Figure 5. Cumulative records for participants in the 75% accuracy group. The solid data path is cumulative correct responses, the dashed line represents perfect responding, and the shaded region shows the range of performances for participants in the comparison group.

Figure 6. Number of trials before mastery criterion of 15 consecutive correct responses was met displayed by group. Each data point represents one participant. Data points below the delta on the Y-axis, in the shaded region, represent mastery before the phase change from inaccurate feedback to accurate feedback.

Figure 7. Percent correct responses during the inaccurate feedback condition by group. Each data point represents one participant. The dotted horizontal line at 0.2 represents chance level responding. The solid horizontal lines display the arithmetic mean for each group.

Figure 8. Percent correct responses during the accurate feedback condition by group. Each data point represents one participant. The dotted horizontal line at 0.2 represents chance level responding. The solid horizontal lines display the arithmetic mean for each group.

Figure 9. Standardized percent area under the curve during the inaccurate feedback condition by group. Solid horizontal lines are the arithmetic mean for each group.

Figure 10. Standardized percent area under the curve during the accurate feedback condition by group. Solid horizontal lines are the arithmetic mean for each group.

Figure 11. Index of strength of incorrect discriminations as represented by the proportion of incorrect selections of comparison stimuli in the presence of all sample stimuli for all participants by group.

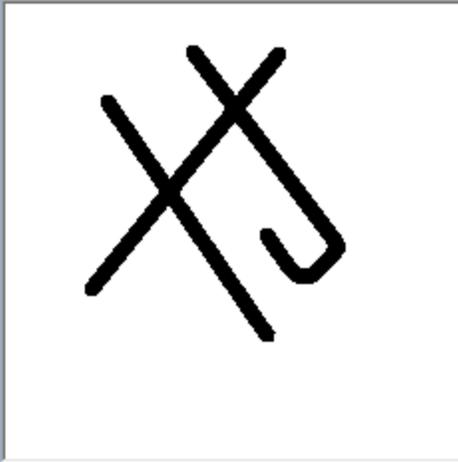
Figure 12. Data for participant A.

Figure 13. Data for participant B.

Figure 14. Data for participant C.

Figure 15. Data for participant D.

Figure 16. Index of strength of incorrect discriminations as represented by the proportion of incorrect selections of comparison stimuli in the presence of each sample stimulus for all participants by condition (top panel) and by task (bottom panel).



Correct

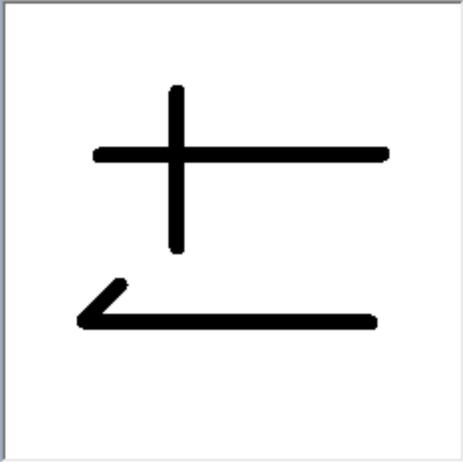
Ipsyn

Punfi

Utzka

Bifdo

Raopol



Incorrect

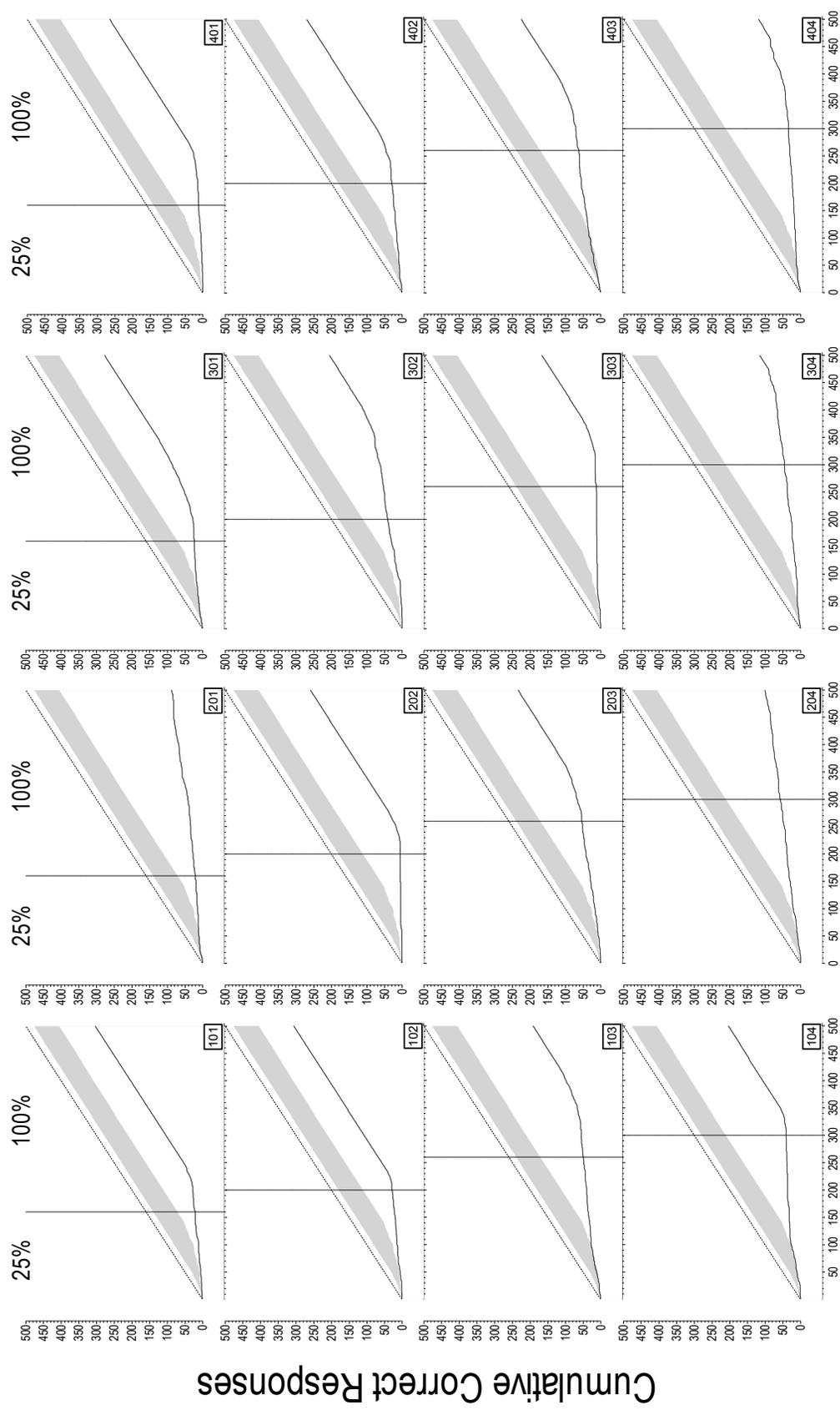
Raopol

Zitaaf

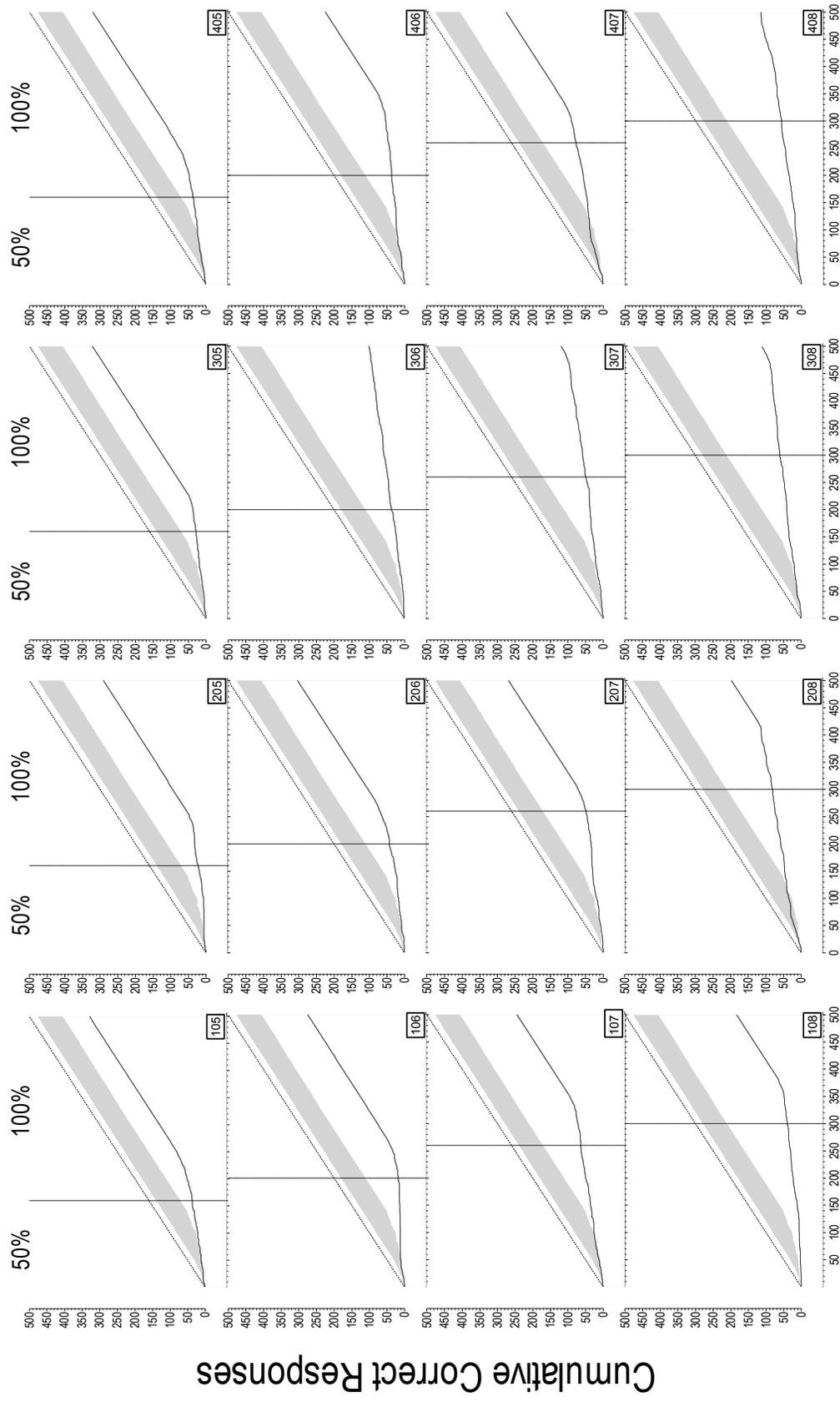
Punfi

Glacko

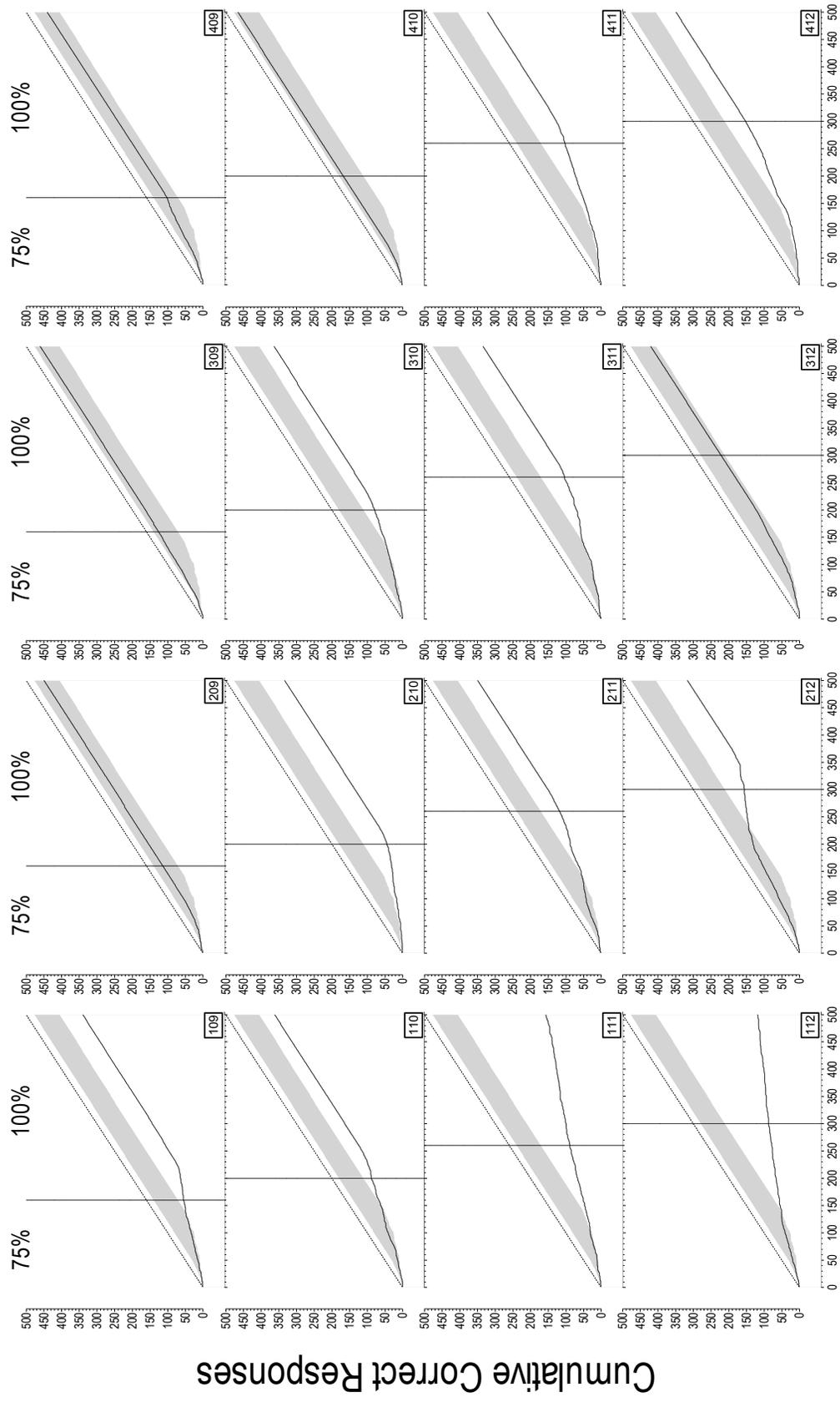
Ipsyn



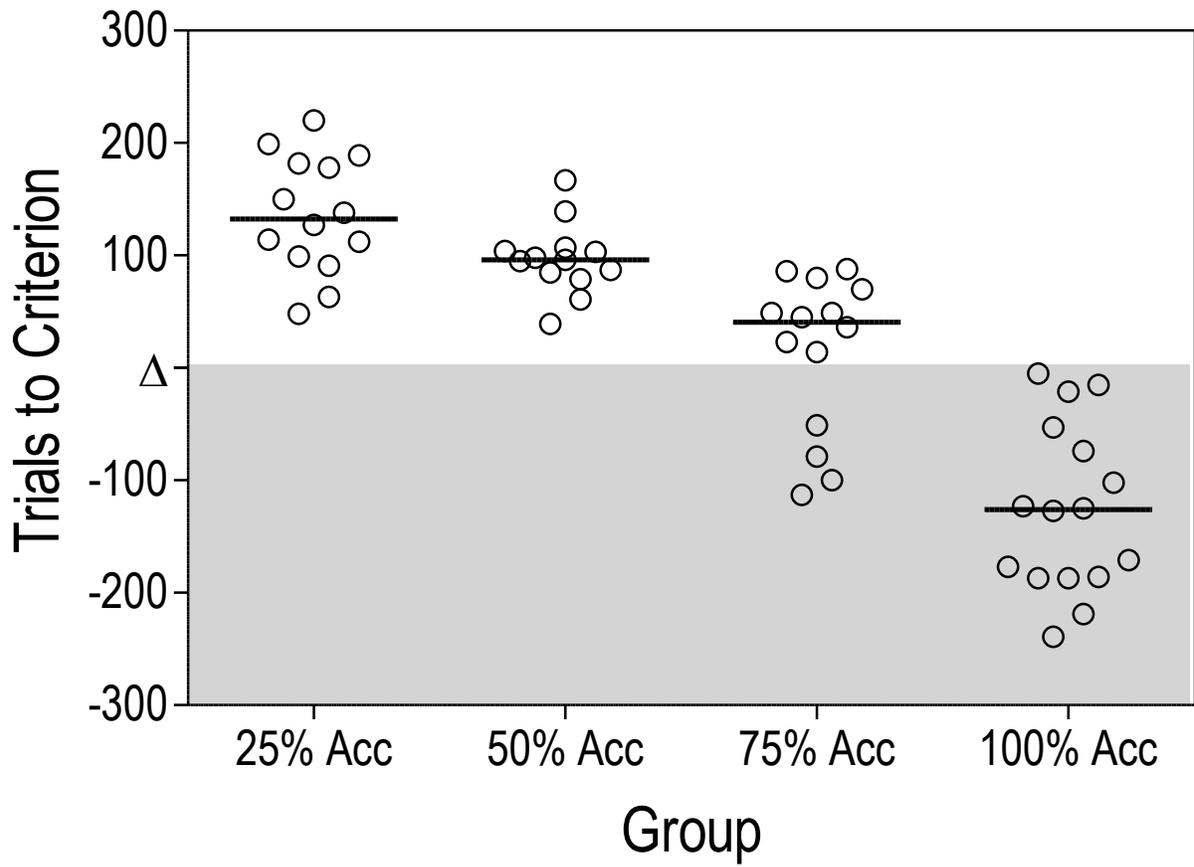
Trials

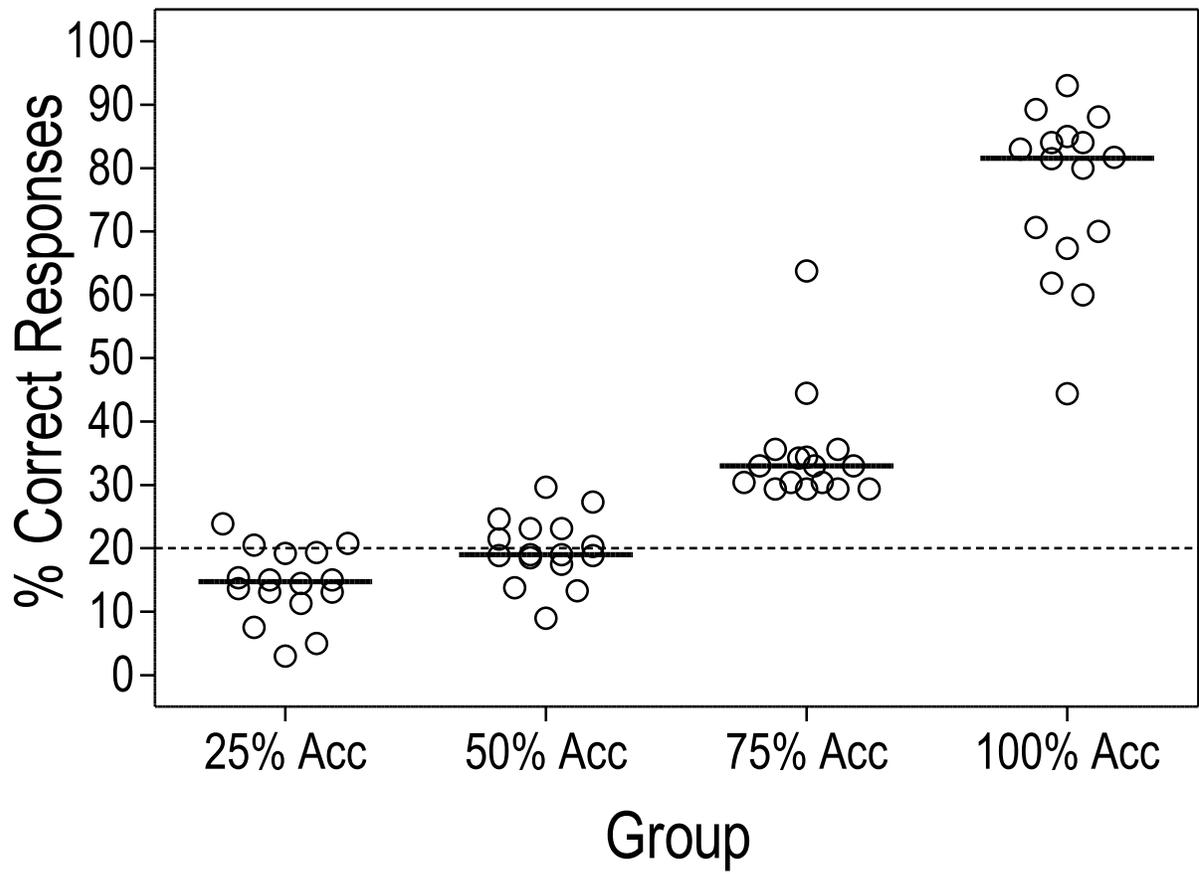


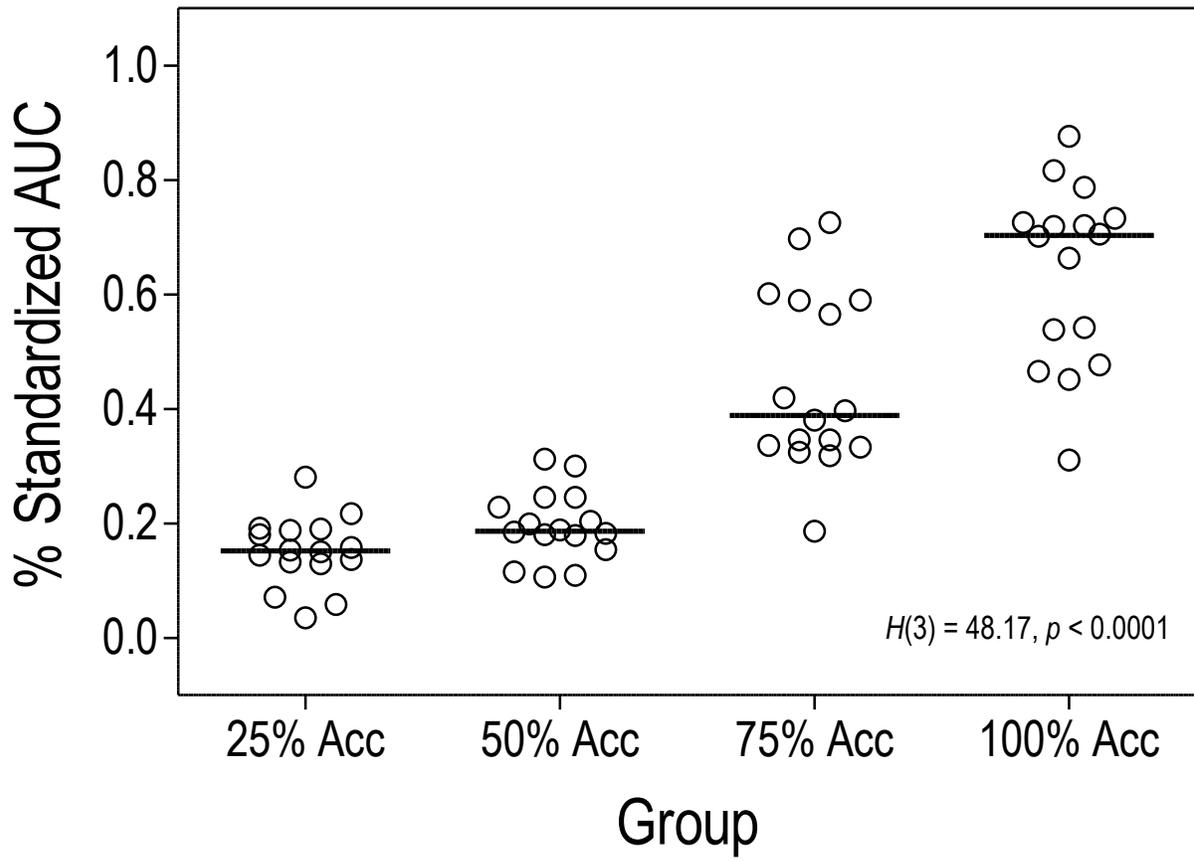
Trials

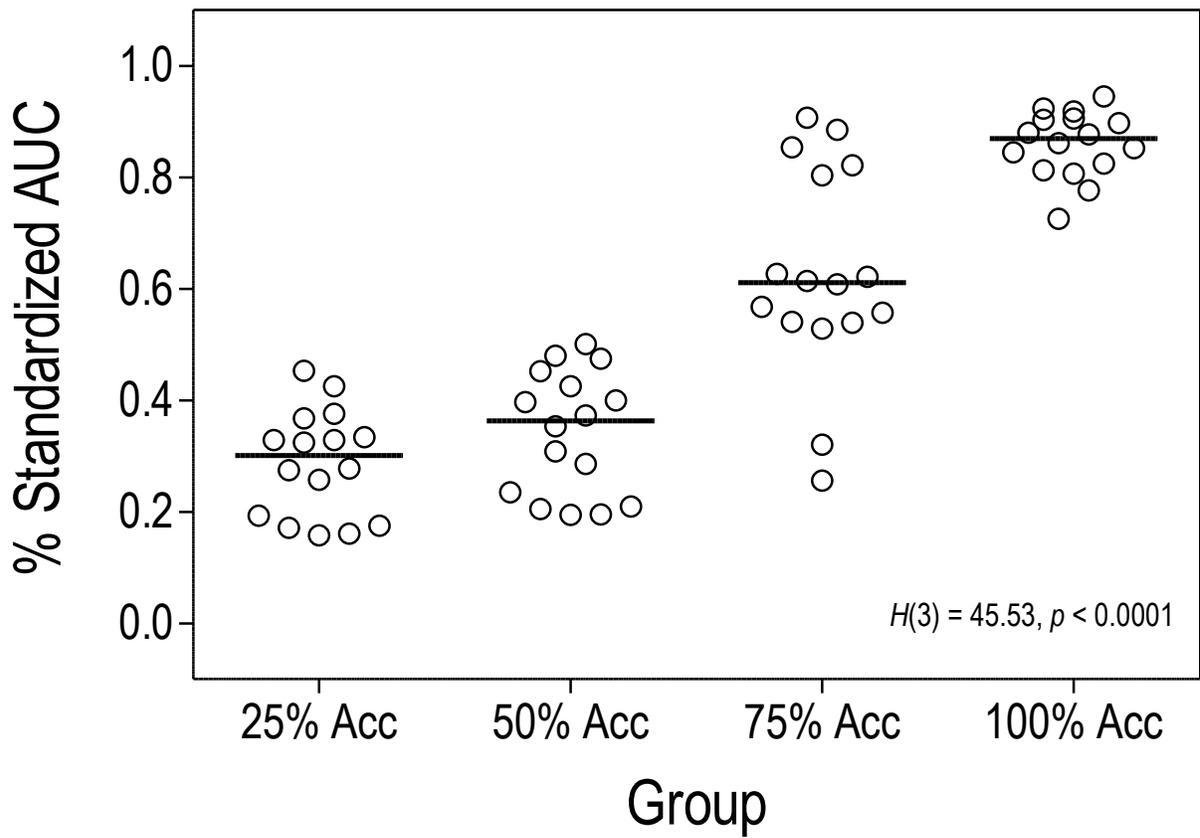


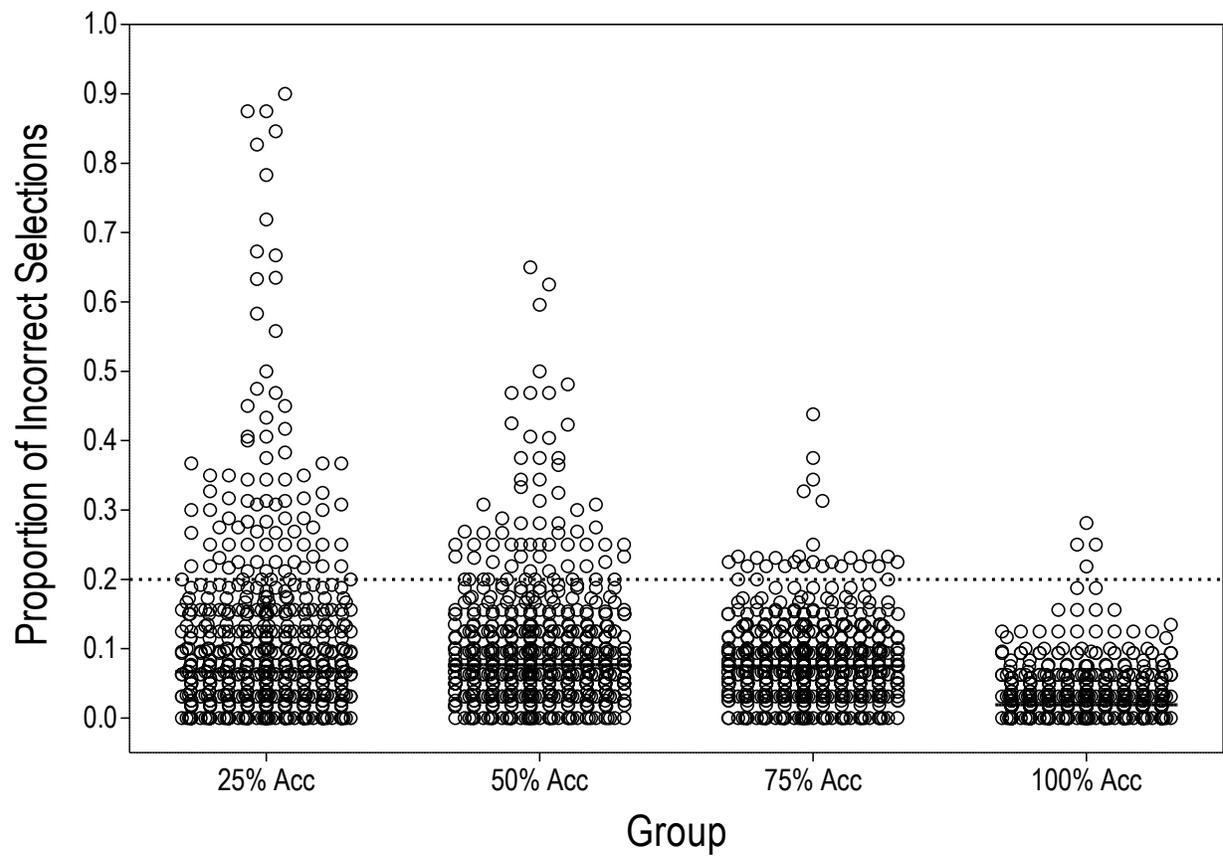
Trials

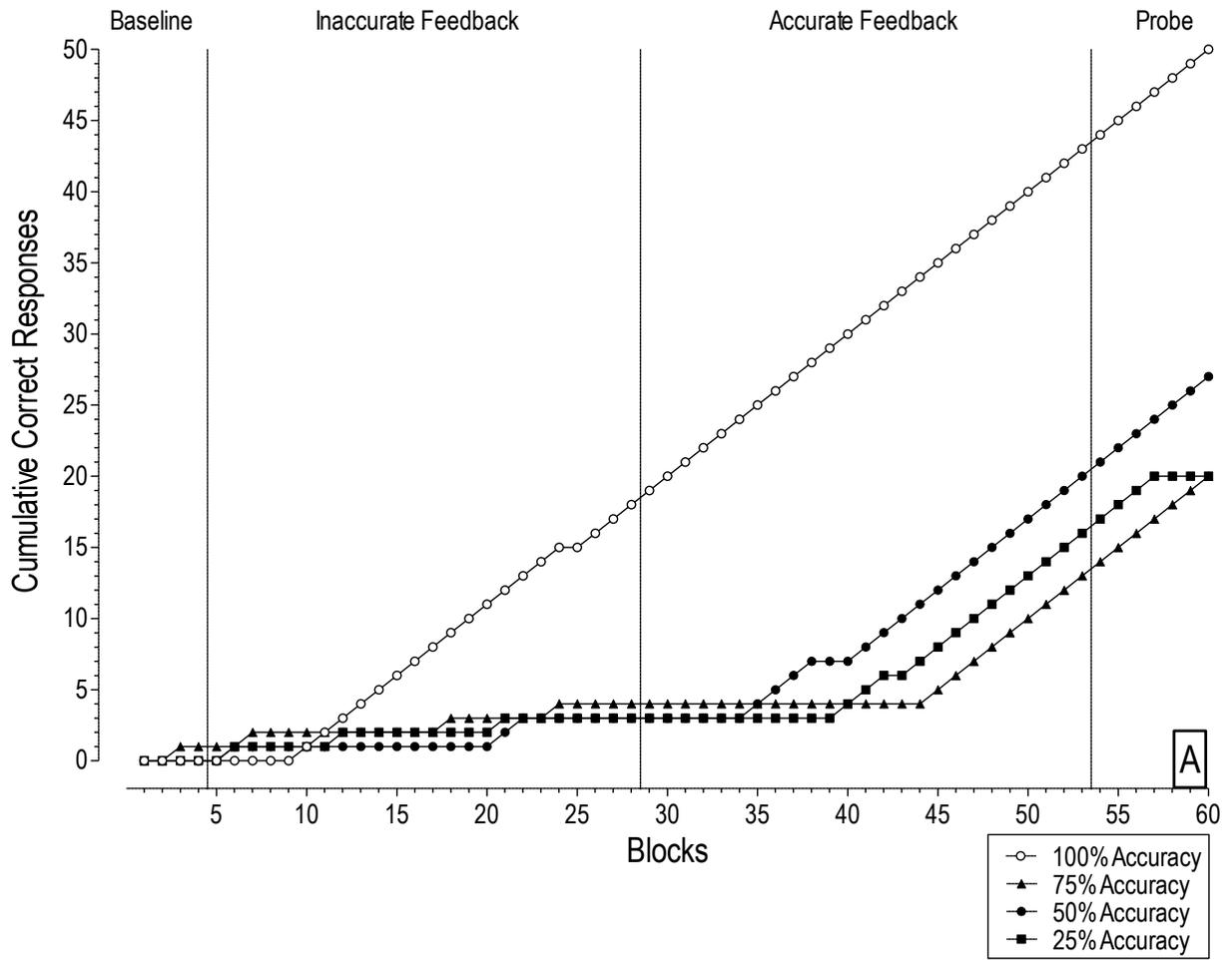


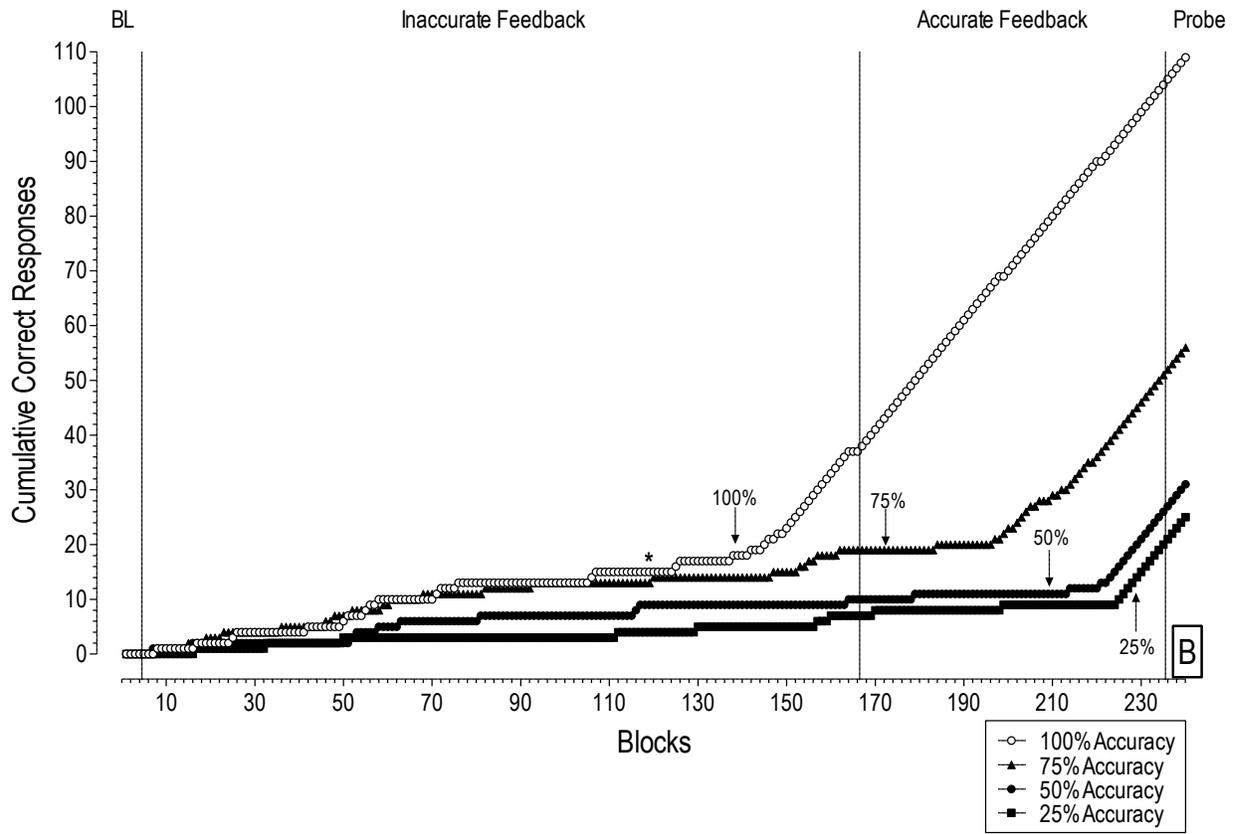


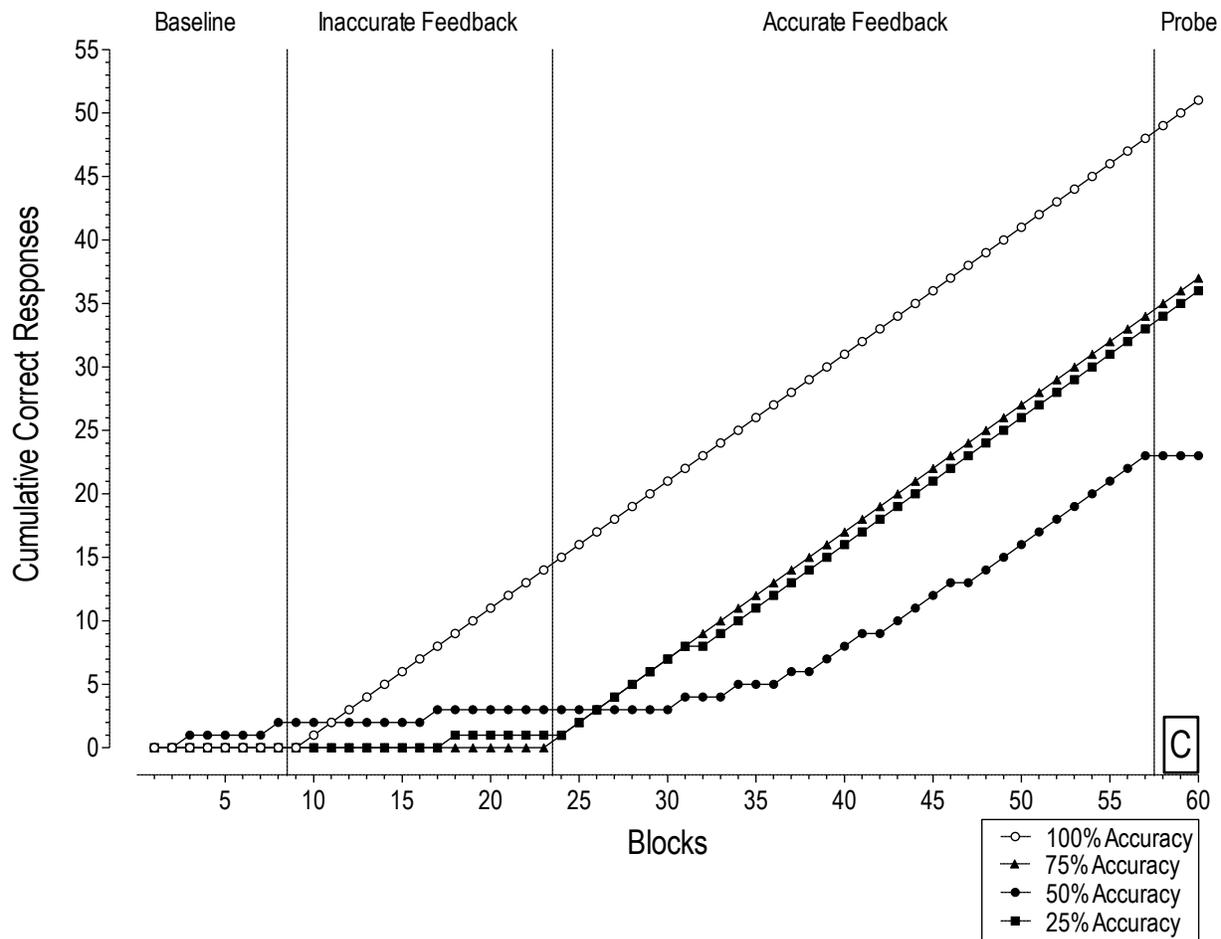


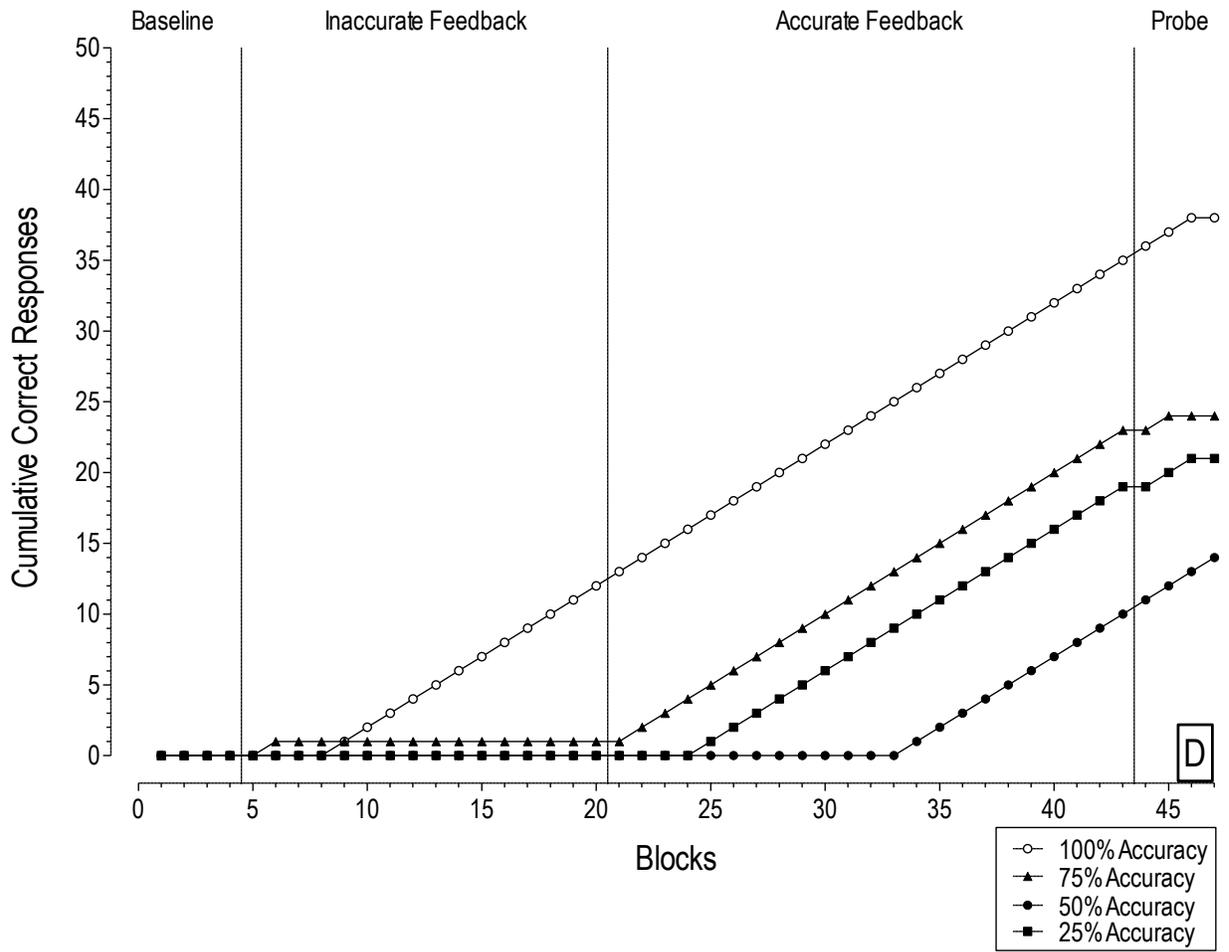


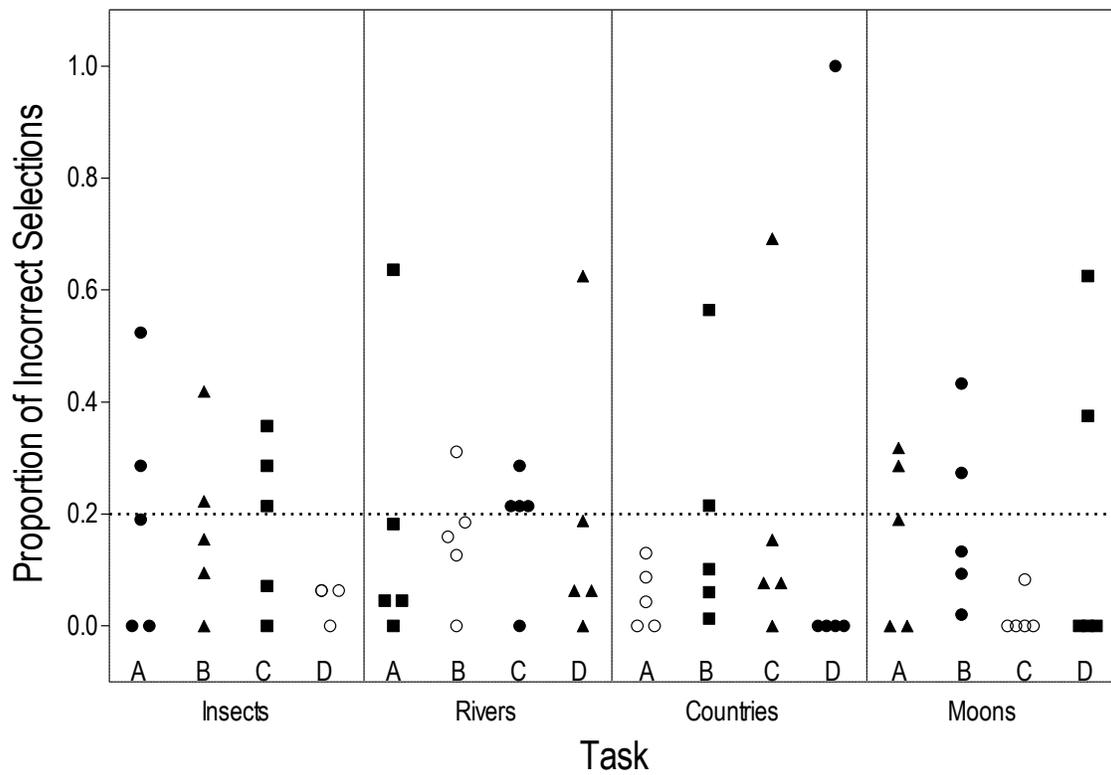
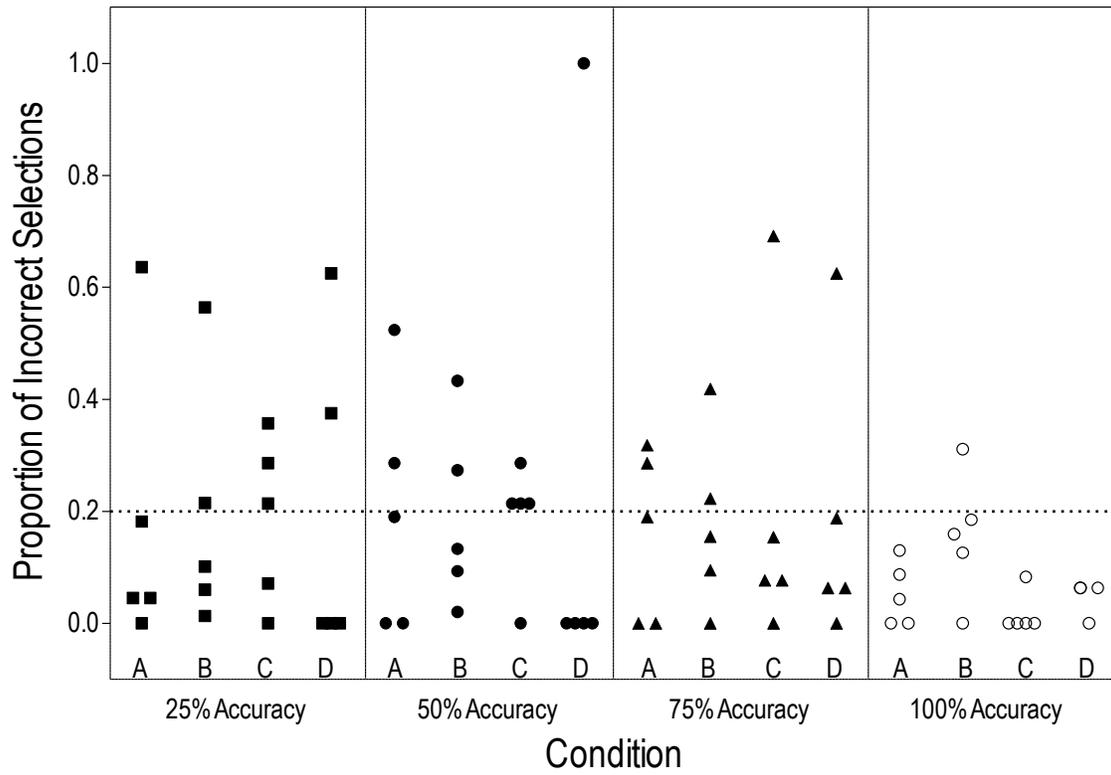












Appendix A

Participant recruitment script

“The purpose of the study is to assess learning under different qualities of instruction. The study will consist of a task presented by a computer program. The task will require that you look at a nonsense shape and choose a nonsense name that corresponds to that shape. The study requires approximately one hour to complete and in exchange for your participation, you will receive 0.5% extra credit towards your final grade in the course, which is about 4 points. You will be able to sign up for a time to participate on the Blackboard site under the Other Extra Credit section in the Research Participation folder. There are currently X sessions scheduled: [*session times*]. You may only participate once in this study so please only sign up for one session. If you sign up and are unable to attend, please remove your name from the sign-up list. If you do attend a session, please arrive on time so that we may start, and end, on time. If anyone has any questions, you may ask me after class or send me an email.”

Appendix B

Demographic survey



The image shows a screenshot of a web-based survey window. The window has a title bar that says "Survey" and a close button in the top right corner. The main content area is light gray and contains the following elements:

- Header text: "Please answer the following:"
- Question 1: "1. Gender" with a dropdown menu.
- Question 2: "2. Age" with a text input field.
- Question 3: "3. Ethnic Background" with a dropdown menu.
- Question 4: "Disability Status" with a dropdown menu.
- Question 5: "Have you had prior exposure to character-based written languages (e.g., Chinese, Japanese, etc.)?" with a dropdown menu.
- A "Submit" button at the bottom center.

Appendix C

Participant Introduction Script

“Welcome and thank you for keeping your appointment. Please read over this consent form and if you agree to participate, please sign the last page and retain the first page for your records. I will be around to collect them when you have finished.”

<<after collecting consent forms, assign participants to a computer>>

“Thank you for your participation in this study. The purpose of the task you will complete is to assess learning under different qualities of teaching. Prior to beginning the task, you will be asked to complete a brief survey. Please note that the information you give on this survey will be kept confidential and will not be linked to your identity in any way. Your identity will not be revealed at any time without your explicit written consent and your performance on the task will be linked only to a participant number. If you have a question during the course of the study, please remain seated and raise your hand. The experimenter will be available to assist you. To maintain conditions required for the study, some questions will not be answered until all participants have completed the study. The experimenter will record your question and will follow-up via email at the conclusion of data collection.”

Appendix D

Indices of Discrimination Strength

25% Accuracy Group 1

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
101	zitaaf	0.219	0.063	0.063	0.063	0.188	0.094	0.125	0.031	0.156
	raopol	0.031	0.094	0.125	0.031	0.156	0.094	0.156	0.156	0.156
	smuzy	0.063	0.219	0.094	0.063	0.219	0.031	0.063	0.094	0.156
	punfi	0.156	0.063	0.094	0.156	0.063	0.125	0.094	0.063	0.188
	bifdo	0.156	0.063	0.063	0.000	0.094	0.156	0.094	0.156	0.219

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
102	zitaaf	0.150	0.025	0.050	0.000	0.275	0.050	0.100	0.200	0.150
	raopol	0.050	0.200	0.125	0.050	0.250	0.050	0.075	0.050	0.150
	smuzy	0.050	0.025	0.100	0.025	0.250	0.175	0.150	0.075	0.150
	punfi	0.025	0.075	0.050	0.125	0.250	0.050	0.150	0.050	0.225
	bifdo	0.350	0.050	0.025	0.025	0.150	0.025	0.025	0.225	0.125

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
103	zitaaf	0.327	0.077	0.077	0.135	0.288	0.019	0.038	0.019	0.019
	raopol	0.096	0.019	0.000	0.000	0.308	0.019	0.000	0.000	0.558
	smuzy	0.038	0.000	0.019	0.019	0.827	0.000	0.000	0.000	0.096
	punfi	0.038	0.269	0.115	0.519	0.019	0.000	0.019	0.019	0.000
	bifdo	0.846	0.000	0.000	0.000	0.077	0.077	0.000	0.000	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
104	zitaaf	0.067	0.017	0.100	0.033	0.117	0.033	0.167	0.367	0.100
	raopol	0.017	0.017	0.000	0.000	0.067	0.050	0.633	0.017	0.200
	smuzy	0.000	0.067	0.400	0.167	0.017	0.183	0.000	0.083	0.083
	punfi	0.067	0.017	0.000	0.050	0.783	0.000	0.000	0.067	0.017
	bifdo	0.083	0.017	0.000	0.033	0.150	0.150	0.250	0.317	0.000

25% Accuracy Group 2

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
201	zitaaf	0.344	0.031	0.156	0.125	0.063	0.063	0.063	0.031	0.125
	raopol	0.031	0.031	0.125	0.000	0.031	0.250	0.156	0.031	0.344
	smuzy	0.156	0.000	0.063	0.000	0.406	0.188	0.031	0.000	0.156
	punfi	0.063	0.000	0.156	0.156	0.094	0.031	0.125	0.031	0.344
	bifdo	0.469	0.000	0.031	0.031	0.063	0.188	0.156	0.000	0.063

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
202	zitaaf	0.025	0.025	0.000	0.025	0.900	0.000	0.000	0.000	0.025
	raopol	0.025	0.025	0.000	0.000	0.000	0.450	0.200	0.000	0.300
	smuzy	0.000	0.150	0.000	0.050	0.075	0.075	0.250	0.000	0.400
	punfi	0.025	0.025	0.000	0.025	0.025	0.200	0.375	0.000	0.325
	bifdo	0.875	0.000	0.000	0.025	0.075	0.025	0.000	0.000	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
203	zitaaf	0.288	0.058	0.115	0.038	0.212	0.058	0.154	0.058	0.019
	raopol	0.058	0.231	0.058	0.038	0.115	0.019	0.096	0.096	0.288
	smuzy	0.115	0.115	0.192	0.096	0.115	0.077	0.173	0.019	0.096
	punfi	0.038	0.019	0.000	0.212	0.115	0.058	0.115	0.308	0.135
	bifdo	0.173	0.058	0.038	0.000	0.115	0.288	0.115	0.154	0.058

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
204	zitaaf	0.133	0.050	0.033	0.067	0.133	0.183	0.133	0.100	0.167
	raopol	0.067	0.217	0.067	0.033	0.233	0.150	0.133	0.050	0.050
	smuzy	0.033	0.033	0.283	0.067	0.133	0.067	0.133	0.083	0.167
	punfi	0.067	0.083	0.033	0.250	0.267	0.017	0.117	0.100	0.067
	bifdo	0.167	0.067	0.150	0.067	0.083	0.100	0.033	0.050	0.283

25% Accuracy Group 3

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
301	zitaaf	0.313	0.000	0.125	0.000	0.000	0.125	0.094	0.344	0.000
	raopol	0.156	0.063	0.031	0.188	0.125	0.094	0.156	0.188	0.000
	smuzy	0.063	0.000	0.344	0.156	0.000	0.125	0.188	0.094	0.031
	punfi	0.000	0.000	0.031	0.031	0.875	0.000	0.063	0.000	0.000
	bifdo	0.406	0.000	0.000	0.219	0.000	0.156	0.156	0.063	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
302	zitaaf	0.275	0.050	0.075	0.025	0.150	0.125	0.100	0.125	0.075
	raopol	0.100	0.200	0.025	0.025	0.075	0.025	0.150	0.225	0.175
	smuzy	0.100	0.075	0.225	0.050	0.175	0.075	0.100	0.125	0.075
	punfi	0.075	0.050	0.000	0.275	0.125	0.075	0.075	0.200	0.125
	bifdo	0.300	0.000	0.025	0.050	0.050	0.125	0.175	0.150	0.125

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
303	zitaaf	0.096	0.000	0.192	0.192	0.077	0.058	0.096	0.096	0.192
	raopol	0.096	0.000	0.038	0.058	0.635	0.038	0.038	0.058	0.038
	smuzy	0.019	0.038	0.077	0.019	0.673	0.038	0.019	0.096	0.019
	punfi	0.192	0.000	0.019	0.019	0.000	0.269	0.173	0.308	0.019
	bifdo	0.058	0.019	0.212	0.192	0.058	0.000	0.192	0.096	0.173

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
304	zitaaf	0.100	0.100	0.117	0.067	0.083	0.150	0.217	0.083	0.083
	raopol	0.217	0.150	0.050	0.050	0.067	0.083	0.033	0.317	0.033
	smuzy	0.000	0.017	0.283	0.033	0.367	0.133	0.050	0.083	0.033
	punfi	0.050	0.100	0.017	0.200	0.450	0.050	0.050	0.050	0.033
	bifdo	0.667	0.050	0.033	0.033	0.033	0.033	0.000	0.133	0.017

25% Accuracy Group 4

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
401	zitaaf	0.188	0.063	0.031	0.063	0.156	0.156	0.188	0.031	0.125
	raopol	0.094	0.031	0.000	0.031	0.250	0.188	0.313	0.000	0.094
	smuzy	0.063	0.031	0.063	0.000	0.313	0.125	0.313	0.000	0.094
	punfi	0.094	0.031	0.000	0.000	0.500	0.063	0.125	0.063	0.125
	bifdo	0.719	0.000	0.000	0.031	0.094	0.031	0.094	0.031	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
402	zitaaf	0.300	0.025	0.050	0.025	0.300	0.100	0.100	0.025	0.075
	raopol	0.125	0.050	0.050	0.050	0.250	0.075	0.225	0.100	0.075
	smuzy	0.100	0.050	0.100	0.025	0.350	0.075	0.150	0.050	0.100
	punfi	0.075	0.050	0.000	0.150	0.475	0.075	0.150	0.000	0.025
	bifdo	0.275	0.075	0.025	0.025	0.150	0.100	0.275	0.050	0.025

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
403	zitaaf	0.327	0.058	0.038	0.038	0.135	0.077	0.173	0.038	0.115
	raopol	0.058	0.250	0.038	0.058	0.154	0.077	0.115	0.019	0.231
	smuzy	0.096	0.058	0.231	0.038	0.135	0.212	0.077	0.077	0.077
	punfi	0.077	0.077	0.173	0.250	0.173	0.077	0.096	0.019	0.058
	bifdo	0.135	0.000	0.019	0.115	0.135	0.019	0.154	0.327	0.096

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
404	zitaaf	0.033	0.033	0.033	0.033	0.350	0.033	0.433	0.000	0.050
	raopol	0.367	0.050	0.133	0.267	0.100	0.000	0.067	0.017	0.000
	smuzy	0.000	0.033	0.117	0.050	0.200	0.000	0.017	0.000	0.583
	punfi	0.000	0.000	0.050	0.067	0.017	0.417	0.067	0.383	0.000
	bifdo	0.050	0.283	0.083	0.083	0.300	0.083	0.100	0.000	0.017

50% Accuracy Group 1

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
105	zitaaf	0.375	0.000	0.063	0.031	0.125	0.000	0.031	0.281	0.094
	raopol	0.094	0.125	0.250	0.063	0.125	0.063	0.156	0.000	0.125
	smuzy	0.000	0.063	0.344	0.063	0.125	0.000	0.031	0.000	0.375
	punfi	0.375	0.000	0.094	0.031	0.031	0.031	0.031	0.344	0.063
	bifdo	0.344	0.031	0.125	0.031	0.281	0.000	0.031	0.063	0.094

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
106	zitaaf	0.075	0.200	0.175	0.125	0.050	0.175	0.050	0.050	0.100
	raopol	0.000	0.050	0.050	0.075	0.625	0.000	0.000	0.000	0.200
	smuzy	0.000	0.000	0.150	0.025	0.650	0.000	0.025	0.075	0.075
	punfi	0.075	0.050	0.100	0.125	0.000	0.125	0.300	0.175	0.050
	bifdo	0.125	0.150	0.125	0.125	0.050	0.150	0.050	0.150	0.075

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
107	zitaaf	0.250	0.058	0.058	0.096	0.096	0.096	0.135	0.058	0.154
	raopol	0.000	0.135	0.096	0.038	0.135	0.058	0.038	0.365	0.135
	smuzy	0.019	0.019	0.154	0.019	0.038	0.154	0.038	0.077	0.481
	punfi	0.250	0.173	0.115	0.212	0.058	0.038	0.115	0.038	0.000
	bifdo	0.077	0.019	0.038	0.038	0.481	0.058	0.173	0.058	0.058

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
108	zitaaf	0.133	0.100	0.067	0.050	0.167	0.033	0.050	0.167	0.233
	raopol	0.050	0.117	0.050	0.100	0.267	0.067	0.150	0.117	0.083
	smuzy	0.033	0.050	0.133	0.100	0.333	0.150	0.083	0.033	0.083
	punfi	0.067	0.033	0.067	0.167	0.100	0.183	0.083	0.183	0.117
	bifdo	0.250	0.067	0.067	0.100	0.117	0.050	0.200	0.033	0.117

50% Accuracy Group 2

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
205	zitaaf	0.344	0.125	0.063	0.031	0.094	0.063	0.188	0.063	0.031
	raopol	0.000	0.031	0.000	0.000	0.406	0.031	0.031	0.031	0.469
	smuzy	0.031	0.219	0.188	0.156	0.000	0.156	0.094	0.156	0.000
	punfi	0.063	0.000	0.063	0.031	0.469	0.063	0.094	0.031	0.188
	bifdo	0.094	0.000	0.094	0.188	0.094	0.063	0.219	0.188	0.063

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
206	zitaaf	0.100	0.075	0.025	0.150	0.250	0.100	0.150	0.100	0.050
	raopol	0.125	0.200	0.100	0.025	0.050	0.075	0.150	0.250	0.025
	smuzy	0.000	0.025	0.400	0.000	0.375	0.025	0.050	0.050	0.075
	punfi	0.050	0.075	0.000	0.150	0.075	0.325	0.150	0.050	0.125
	bifdo	0.175	0.050	0.050	0.050	0.225	0.050	0.075	0.250	0.075

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
207	zitaaf	0.365	0.019	0.058	0.077	0.135	0.096	0.058	0.115	0.077
	raopol	0.077	0.135	0.077	0.077	0.308	0.058	0.077	0.096	0.096
	smuzy	0.096	0.077	0.173	0.058	0.154	0.135	0.135	0.038	0.135
	punfi	0.115	0.058	0.096	0.096	0.288	0.096	0.115	0.115	0.019
	bifdo	0.231	0.077	0.077	0.077	0.173	0.115	0.058	0.096	0.096

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
208	zitaaf	0.217	0.083	0.017	0.050	0.133	0.100	0.267	0.050	0.083
	raopol	0.100	0.150	0.083	0.017	0.200	0.100	0.100	0.117	0.133
	smuzy	0.017	0.033	0.317	0.083	0.117	0.000	0.067	0.233	0.133
	punfi	0.117	0.150	0.033	0.367	0.117	0.067	0.033	0.017	0.100
	bifdo	0.200	0.083	0.050	0.033	0.317	0.100	0.050	0.067	0.100

50% Accuracy Group 3

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
305	zitaaf	0.156	0.031	0.000	0.000	0.500	0.125	0.063	0.063	0.063
	raopol	0.031	0.094	0.031	0.063	0.125	0.000	0.156	0.469	0.031
	smuzy	0.063	0.094	0.188	0.063	0.281	0.031	0.063	0.063	0.156
	punfi	0.031	0.156	0.063	0.250	0.188	0.000	0.063	0.125	0.125
	bifdo	0.188	0.063	0.094	0.031	0.188	0.156	0.031	0.063	0.188

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
306	zitaaf	0.200	0.050	0.075	0.050	0.225	0.075	0.100	0.100	0.125
	raopol	0.025	0.225	0.050	0.025	0.200	0.075	0.050	0.275	0.075
	smuzy	0.050	0.050	0.125	0.075	0.100	0.225	0.250	0.050	0.075
	punfi	0.050	0.025	0.025	0.250	0.250	0.125	0.050	0.125	0.100
	bifdo	0.425	0.000	0.075	0.025	0.150	0.125	0.025	0.050	0.125

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
307	zitaaf	0.154	0.115	0.038	0.192	0.192	0.077	0.154	0.019	0.058
	raopol	0.058	0.135	0.154	0.077	0.269	0.096	0.096	0.038	0.077
	smuzy	0.038	0.019	0.250	0.096	0.173	0.038	0.096	0.135	0.154
	punfi	0.077	0.173	0.058	0.250	0.154	0.058	0.077	0.077	0.077
	bifdo	0.212	0.038	0.019	0.019	0.154	0.096	0.096	0.269	0.096

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
308	zitaaf	0.217	0.067	0.133	0.067	0.100	0.100	0.050	0.100	0.167
	raopol	0.083	0.200	0.100	0.050	0.133	0.100	0.133	0.033	0.167
	smuzy	0.150	0.050	0.200	0.033	0.133	0.067	0.150	0.050	0.167
	punfi	0.050	0.050	0.083	0.167	0.133	0.133	0.083	0.233	0.067
	bifdo	0.250	0.017	0.033	0.033	0.233	0.167	0.067	0.117	0.083

50% Accuracy Group 4

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
405	zitaaf	0.500	0.000	0.125	0.063	0.094	0.031	0.063	0.031	0.094
	raopol	0.031	0.188	0.156	0.125	0.281	0.031	0.000	0.094	0.094
	smuzy	0.219	0.094	0.125	0.063	0.063	0.031	0.313	0.000	0.094
	punfi	0.188	0.125	0.063	0.063	0.094	0.250	0.063	0.031	0.125
	bifdo	0.125	0.031	0.063	0.125	0.281	0.125	0.063	0.094	0.094

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
406	zitaaf	0.225	0.075	0.075	0.125	0.250	0.050	0.050	0.075	0.075
	raopol	0.100	0.200	0.050	0.000	0.125	0.125	0.200	0.075	0.125
	smuzy	0.050	0.025	0.175	0.025	0.200	0.050	0.200	0.150	0.125
	punfi	0.100	0.025	0.075	0.200	0.175	0.150	0.125	0.075	0.075
	bifdo	0.175	0.125	0.050	0.100	0.125	0.150	0.025	0.100	0.150

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
407	zitaaf	0.519	0.038	0.077	0.038	0.192	0.019	0.019	0.077	0.019
	raopol	0.038	0.038	0.058	0.000	0.058	0.058	0.308	0.038	0.404
	smuzy	0.000	0.038	0.231	0.000	0.596	0.077	0.019	0.019	0.019
	punfi	0.038	0.212	0.019	0.673	0.000	0.000	0.019	0.019	0.019
	bifdo	0.096	0.000	0.077	0.019	0.019	0.115	0.423	0.000	0.250

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist. 1	Dist. 2	Dist. 3	Dist. 4
408	zitaaf	0.317	0.050	0.083	0.083	0.167	0.117	0.083	0.050	0.050
	raopol	0.050	0.183	0.033	0.100	0.267	0.167	0.050	0.033	0.117
	smuzy	0.050	0.050	0.200	0.133	0.233	0.083	0.000	0.150	0.100
	punfi	0.083	0.017	0.067	0.167	0.200	0.117	0.117	0.100	0.133
	bifdo	0.200	0.117	0.050	0.183	0.083	0.133	0.050	0.067	0.117

75% Accuracy Group 1

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
109	zitaaf	0.094	0.000	0.000	0.094	0.219	0.125	0.000	0.031	0.438
	raopol	0.000	0.781	0.031	0.000	0.031	0.063	0.031	0.000	0.063
	smuzy	0.000	0.031	0.063	0.031	0.219	0.250	0.125	0.063	0.219
	punfi	0.094	0.031	0.031	0.500	0.188	0.063	0.000	0.031	0.063
	bifdo	0.094	0.063	0.031	0.156	0.281	0.219	0.031	0.031	0.094

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
110	zitaaf	0.500	0.125	0.075	0.025	0.075	0.025	0.050	0.075	0.050
	raopol	0.000	0.425	0.000	0.025	0.375	0.025	0.050	0.050	0.050
	smuzy	0.025	0.050	0.700	0.025	0.050	0.025	0.050	0.050	0.025
	punfi	0.050	0.075	0.000	0.475	0.050	0.150	0.025	0.050	0.125
	bifdo	0.225	0.075	0.125	0.050	0.125	0.050	0.000	0.225	0.125

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
111	zitaaf	0.481	0.077	0.038	0.000	0.115	0.038	0.019	0.135	0.096
	raopol	0.058	0.173	0.096	0.000	0.058	0.077	0.115	0.096	0.327
	smuzy	0.096	0.096	0.038	0.058	0.135	0.115	0.135	0.173	0.154
	punfi	0.096	0.096	0.096	0.250	0.192	0.077	0.038	0.096	0.058
	bifdo	0.019	0.038	0.019	0.019	0.769	0.058	0.019	0.038	0.019

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
112	zitaaf	0.267	0.050	0.083	0.067	0.067	0.167	0.083	0.067	0.150
	raopol	0.050	0.500	0.050	0.050	0.100	0.067	0.033	0.050	0.100
	smuzy	0.083	0.117	0.300	0.083	0.117	0.083	0.067	0.100	0.050
	punfi	0.050	0.050	0.033	0.300	0.233	0.067	0.050	0.100	0.117
	bifdo	0.167	0.067	0.133	0.083	0.100	0.133	0.067	0.100	0.150

75%Accuracy Group 2

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
209	zitaaf	0.313	0.031	0.125	0.094	0.031	0.125	0.094	0.094	0.094
	raopol	0.000	0.625	0.000	0.031	0.094	0.063	0.031	0.063	0.094
	smuzy	0.125	0.031	0.406	0.094	0.156	0.031	0.000	0.094	0.063
	punfi	0.031	0.031	0.000	0.438	0.219	0.094	0.063	0.031	0.094
	bifdo	0.188	0.031	0.188	0.156	0.000	0.219	0.031	0.031	0.156

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
210	zitaaf	0.325	0.050	0.100	0.075	0.050	0.125	0.100	0.075	0.100
	raopol	0.000	0.550	0.050	0.075	0.100	0.050	0.025	0.050	0.100
	smuzy	0.100	0.025	0.350	0.125	0.150	0.075	0.025	0.100	0.050
	punfi	0.075	0.050	0.000	0.400	0.225	0.075	0.050	0.050	0.075
	bifdo	0.200	0.050	0.150	0.125	0.025	0.175	0.075	0.050	0.150

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
211	zitaaf	0.308	0.058	0.077	0.058	0.077	0.135	0.096	0.077	0.115
	raopol	0.038	0.500	0.058	0.058	0.115	0.058	0.019	0.038	0.115
	smuzy	0.096	0.077	0.308	0.096	0.135	0.096	0.058	0.077	0.058
	punfi	0.058	0.038	0.000	0.308	0.231	0.077	0.058	0.096	0.135
	bifdo	0.173	0.077	0.135	0.096	0.096	0.154	0.058	0.077	0.135

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
212	zitaaf	0.267	0.050	0.083	0.067	0.067	0.167	0.083	0.067	0.150
	raopol	0.050	0.500	0.050	0.050	0.100	0.067	0.033	0.050	0.100
	smuzy	0.083	0.117	0.300	0.083	0.117	0.083	0.067	0.100	0.050
	punfi	0.050	0.050	0.033	0.300	0.233	0.067	0.050	0.100	0.117
	bifdo	0.167	0.067	0.133	0.083	0.100	0.133	0.067	0.100	0.150

75% Accuracy Group 3

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
309	zitaaf	0.313	0.031	0.125	0.094	0.031	0.125	0.094	0.094	0.094
	raopol	0.000	0.625	0.000	0.031	0.094	0.063	0.031	0.063	0.094
	smuzy	0.125	0.031	0.406	0.094	0.156	0.031	0.000	0.094	0.063
	punfi	0.031	0.031	0.000	0.438	0.219	0.094	0.063	0.031	0.094
	bifdo	0.188	0.031	0.188	0.156	0.000	0.219	0.031	0.031	0.156

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
310	zitaaf	0.325	0.050	0.100	0.075	0.050	0.125	0.100	0.075	0.100
	raopol	0.000	0.550	0.050	0.075	0.100	0.050	0.025	0.050	0.100
	smuzy	0.100	0.025	0.350	0.125	0.150	0.075	0.025	0.100	0.050
	punfi	0.075	0.050	0.000	0.400	0.225	0.075	0.050	0.050	0.075
	bifdo	0.200	0.050	0.150	0.125	0.025	0.175	0.075	0.050	0.150

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
311	zitaaf	0.308	0.058	0.077	0.058	0.077	0.135	0.096	0.077	0.115
	raopol	0.038	0.500	0.058	0.058	0.115	0.058	0.019	0.038	0.115
	smuzy	0.096	0.077	0.308	0.096	0.135	0.096	0.058	0.077	0.058
	punfi	0.058	0.038	0.000	0.308	0.231	0.077	0.058	0.096	0.135
	bifdo	0.173	0.077	0.135	0.096	0.096	0.154	0.058	0.077	0.135

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
312	zitaaf	0.267	0.050	0.083	0.067	0.067	0.167	0.083	0.067	0.150
	raopol	0.050	0.500	0.050	0.050	0.100	0.067	0.033	0.050	0.100
	smuzy	0.083	0.117	0.300	0.083	0.117	0.083	0.067	0.100	0.050
	punfi	0.050	0.050	0.033	0.300	0.233	0.067	0.050	0.100	0.117
	bifdo	0.167	0.067	0.133	0.083	0.100	0.133	0.067	0.100	0.150

75% Accuracy Group 4

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
409	zitaaf	0.281	0.000	0.125	0.000	0.063	0.063	0.344	0.031	0.094
	raopol	0.000	0.938	0.000	0.000	0.031	0.031	0.000	0.000	0.000
	smuzy	0.000	0.000	0.813	0.000	0.063	0.031	0.000	0.000	0.094
	punfi	0.000	0.000	0.031	0.844	0.031	0.063	0.000	0.031	0.000
	bifdo	0.125	0.000	0.000	0.000	0.313	0.031	0.156	0.063	0.313

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
410	zitaaf	0.325	0.050	0.100	0.075	0.050	0.125	0.100	0.075	0.100
	raopol	0.000	0.550	0.050	0.075	0.100	0.050	0.025	0.050	0.100
	smuzy	0.100	0.025	0.350	0.125	0.150	0.075	0.025	0.100	0.050
	punfi	0.075	0.050	0.000	0.400	0.225	0.075	0.050	0.050	0.075
	bifdo	0.200	0.050	0.150	0.125	0.025	0.175	0.075	0.050	0.150

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
411	zitaaf	0.308	0.058	0.077	0.058	0.077	0.135	0.096	0.077	0.115
	raopol	0.038	0.500	0.058	0.058	0.115	0.058	0.019	0.038	0.115
	smuzy	0.096	0.077	0.308	0.096	0.135	0.096	0.058	0.077	0.058
	punfi	0.058	0.038	0.000	0.308	0.231	0.077	0.058	0.096	0.135
	bifdo	0.173	0.077	0.135	0.096	0.096	0.154	0.058	0.077	0.135

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
412	zitaaf	0.267	0.050	0.083	0.067	0.067	0.167	0.083	0.067	0.150
	raopol	0.050	0.500	0.050	0.050	0.100	0.067	0.033	0.050	0.100
	smuzy	0.083	0.117	0.300	0.083	0.117	0.083	0.067	0.100	0.050
	punfi	0.050	0.050	0.033	0.300	0.233	0.067	0.050	0.100	0.117
	bifdo	0.167	0.067	0.133	0.083	0.100	0.133	0.067	0.100	0.150

100% Accuracy Group 1

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
113	zitaaf	0.469	0.031	0.094	0.000	0.281	0.000	0.063	0.031	0.031
	raopol	0.031	0.625	0.063	0.000	0.063	0.063	0.000	0.094	0.063
	smuzy	0.063	0.000	0.531	0.031	0.000	0.063	0.063	0.156	0.094
	punfi	0.031	0.063	0.000	0.844	0.031	0.000	0.031	0.000	0.000
	bifdo	0.125	0.000	0.000	0.031	0.531	0.094	0.063	0.094	0.063

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
114	zitaaf	0.950	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.025
	raopol	0.000	0.825	0.000	0.025	0.025	0.025	0.075	0.025	0.000
	smuzy	0.000	0.000	0.800	0.000	0.050	0.050	0.025	0.025	0.050
	punfi	0.050	0.000	0.000	0.875	0.025	0.025	0.025	0.000	0.000
	bifdo	0.100	0.000	0.025	0.025	0.700	0.025	0.025	0.050	0.050

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
115	zitaaf	0.885	0.000	0.038	0.000	0.058	0.000	0.000	0.000	0.019
	raopol	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	smuzy	0.000	0.000	0.846	0.000	0.058	0.038	0.019	0.000	0.038
	punfi	0.000	0.000	0.000	0.827	0.058	0.038	0.000	0.000	0.077
	bifdo	0.096	0.000	0.000	0.038	0.846	0.000	0.019	0.000	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
116	zitaaf	0.883	0.000	0.000	0.050	0.017	0.050	0.000	0.000	0.000
	raopol	0.033	0.683	0.017	0.000	0.083	0.033	0.033	0.067	0.050
	smuzy	0.033	0.033	0.833	0.000	0.033	0.050	0.000	0.017	0.000
	punfi	0.050	0.000	0.000	0.917	0.000	0.017	0.000	0.000	0.017
	bifdo	0.083	0.017	0.017	0.017	0.767	0.033	0.017	0.033	0.017

100% Accuracy Group 2

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
213	zitaaf	0.625	0.031	0.031	0.063	0.094	0.063	0.000	0.094	0.000
	raopol	0.000	0.906	0.031	0.031	0.000	0.031	0.000	0.000	0.000
	smuzy	0.000	0.031	0.875	0.000	0.031	0.000	0.063	0.000	0.000
	punfi	0.000	0.031	0.000	0.500	0.063	0.000	0.063	0.125	0.219
	bifdo	0.125	0.000	0.000	0.031	0.625	0.031	0.063	0.094	0.031

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
214	zitaaf	0.750	0.025	0.025	0.000	0.075	0.075	0.025	0.025	0.000
	raopol	0.000	0.875	0.000	0.025	0.000	0.025	0.000	0.025	0.050
	smuzy	0.050	0.000	0.950	0.000	0.000	0.000	0.000	0.000	0.000
	punfi	0.025	0.000	0.000	0.725	0.050	0.050	0.025	0.075	0.050
	bifdo	0.025	0.000	0.000	0.025	0.900	0.025	0.000	0.000	0.025

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
215	zitaaf	0.635	0.058	0.019	0.019	0.115	0.038	0.038	0.019	0.058
	raopol	0.038	0.577	0.038	0.019	0.135	0.077	0.000	0.058	0.058
	smuzy	0.019	0.000	0.808	0.019	0.058	0.038	0.000	0.019	0.038
	punfi	0.000	0.058	0.019	0.712	0.077	0.038	0.038	0.038	0.019
	bifdo	0.058	0.077	0.000	0.038	0.635	0.038	0.096	0.000	0.058

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
216	zitaaf	0.883	0.000	0.017	0.017	0.033	0.017	0.017	0.000	0.017
	raopol	0.000	0.983	0.000	0.000	0.017	0.000	0.000	0.000	0.000
	smuzy	0.033	0.000	0.917	0.000	0.017	0.017	0.017	0.000	0.000
	punfi	0.017	0.000	0.017	0.900	0.000	0.000	0.017	0.017	0.033
	bifdo	0.017	0.000	0.000	0.000	0.967	0.017	0.000	0.000	0.000

100% Accuracy Group 3

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
313	zitaaf	0.375	0.094	0.000	0.063	0.250	0.031	0.031	0.031	0.125
	raopol	0.063	0.375	0.031	0.031	0.094	0.188	0.063	0.094	0.063
	smuzy	0.000	0.031	0.563	0.000	0.156	0.125	0.094	0.000	0.031
	punfi	0.125	0.000	0.031	0.313	0.156	0.156	0.094	0.063	0.063
	bifdo	0.250	0.000	0.063	0.000	0.594	0.063	0.031	0.000	0.000

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
314	zitaaf	0.925	0.000	0.000	0.000	0.025	0.025	0.000	0.025	0.000
	raopol	0.000	0.700	0.100	0.050	0.075	0.000	0.000	0.025	0.050
	smuzy	0.000	0.075	0.525	0.025	0.075	0.050	0.075	0.075	0.100
	punfi	0.075	0.000	0.025	0.700	0.125	0.000	0.025	0.050	0.000
	bifdo	0.025	0.050	0.000	0.025	0.650	0.100	0.050	0.025	0.075

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
315	zitaaf	0.904	0.000	0.019	0.000	0.000	0.038	0.019	0.000	0.019
	raopol	0.000	0.846	0.000	0.038	0.058	0.019	0.019	0.019	0.000
	smuzy	0.000	0.000	0.827	0.000	0.019	0.058	0.019	0.019	0.058
	punfi	0.038	0.019	0.000	0.865	0.000	0.000	0.058	0.000	0.019
	bifdo	0.019	0.000	0.000	0.000	0.808	0.058	0.058	0.038	0.019

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
316	zitaaf	0.750	0.017	0.017	0.050	0.000	0.033	0.050	0.017	0.067
	raopol	0.017	0.783	0.000	0.000	0.117	0.017	0.050	0.000	0.017
	smuzy	0.017	0.017	0.967	0.000	0.000	0.000	0.000	0.000	0.000
	punfi	0.017	0.033	0.000	0.867	0.017	0.017	0.033	0.000	0.017
	bifdo	0.033	0.000	0.000	0.000	0.833	0.033	0.017	0.000	0.083

100% Accuracy Group 4

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
413	zitaaf	0.625	0.031	0.000	0.031	0.094	0.094	0.094	0.031	0.000
	raopol	0.031	0.594	0.031	0.000	0.188	0.063	0.063	0.031	0.000
	smuzy	0.031	0.000	0.875	0.000	0.000	0.000	0.031	0.063	0.000
	punfi	0.063	0.000	0.000	0.531	0.094	0.063	0.125	0.063	0.063
	bifdo	0.125	0.063	0.000	0.063	0.469	0.063	0.094	0.094	0.031

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
414	zitaaf	0.825	0.000	0.000	0.000	0.100	0.000	0.025	0.000	0.050
	raopol	0.000	0.875	0.050	0.025	0.025	0.025	0.000	0.000	0.000
	smuzy	0.025	0.025	0.850	0.000	0.025	0.050	0.025	0.000	0.000
	punfi	0.000	0.000	0.025	0.675	0.000	0.050	0.100	0.075	0.075
	bifdo	0.025	0.000	0.000	0.000	0.850	0.025	0.050	0.025	0.025

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
415	zitaaf	0.942	0.019	0.019	0.000	0.019	0.000	0.000	0.000	0.000
	raopol	0.019	0.923	0.019	0.019	0.000	0.000	0.000	0.000	0.019
	smuzy	0.000	0.019	0.788	0.000	0.058	0.058	0.038	0.000	0.038
	punfi	0.000	0.019	0.000	0.942	0.000	0.000	0.000	0.038	0.000
	bifdo	0.000	0.000	0.019	0.019	0.865	0.058	0.000	0.000	0.038

Participant	Shape	Response								
		zitaaf	raopol	smuzy	punfi	bifdo	Dist 1.	Dist. 2	Dist. 3	Dist. 4
416	zitaaf	0.800	0.000	0.000	0.017	0.033	0.083	0.050	0.017	0.000
	raopol	0.017	0.817	0.000	0.000	0.100	0.017	0.050	0.000	0.000
	smuzy	0.017	0.067	0.700	0.000	0.050	0.050	0.033	0.017	0.067
	punfi	0.033	0.000	0.000	0.917	0.000	0.000	0.017	0.033	0.000
	bifdo	0.050	0.017	0.033	0.033	0.767	0.017	0.033	0.000	0.050

Appendix E

Parent Letter

Dear Parent/Guardian,

My name is Jason Hirst, a graduate student researcher at the University of Kansas, Department of Applied Behavioral Science. As a part of my master's thesis, I am conducting a study assessing how errors made during teaching influence student learning.

If you agree to provide consent for your child to participate, s/he will be taught to complete four academic tasks during the course of the study. The tasks will involve the identification of countries and rivers on a blank outline map, the names of aquatic insects that may be found in local streams, and moons orbiting planets in the solar system. Participants will be able to earn tokens throughout the study, which can be exchanged for rewards. The reward items will be selected based on the results of a preference assessment where your child will be able to choose the items s/he would like to work for and will consist of academically appropriate materials such as school supplies, coloring materials, and stickers. In general, tokens and praise will be delivered following correct responses to the tasks and incorrect responses will be followed by a neutral statement. However, during a portion of the study, the experimenter will make errors during the instruction by giving praise and tokens following incorrect answers with praise and rewards and omitting rewards following correct answers. Previous research has shown that learning is negatively influenced in the presence of teaching errors. The purpose of this study is to determine whether those effects carry-over after errors are corrected.

The study will be conducted during multiple sessions, lasting about 30-60 minutes each. Each session will consist of 30-60 minutes of instruction time with breaks every 15 minutes. I will be working with the school staff to identify a time during the school day that will least interfere with your child's normal educational activities. We estimate that the study will take about 4-6 sessions to complete, making the total time commitment no more than 8 hours.

There are no physical risks associated with participation in this study. Other risks are minimal and may include frustration as a result of the inconsistency of the feedback. It is also possible that your child will learn incorrect information during the imperfect instruction phase. However, the final phase of the study involves giving only accurate feedback and will continue until your child has learned and mastered the correct answers to the tasks, minimizing the chance that s/he will retain incorrect information beyond the end of the study. Should your child feel uncomfortable with participating at any time during the study, s/he may ask to leave without penalty. We will also ask your child if s/he would like to work with us prior to beginning each session. In the case that your child chooses to stop participation prior to learning the correct information, you will be provided with a copy of the materials indicating the correct answers. You may use these materials to provide your child with the correct information if you are concerned about your child retaining false information.

Your confidentiality and the privacy of your child will be protected. Participants will be assigned a number or alias and no real names or identifying information will be released at any time unless required by law. All research materials will be kept in a locked cabinet or on a secure server. If you agree to provide consent, you should know that you have the right to withdraw your consent at any time. The results of this study will contribute to our understanding of how teaching errors impact student learning both in the short- and long-term. If you have any questions or agree to allow your child to participate, please contact Jason Hirst by email at jmhirst@ku.edu.

Appendix F

Informed Consent

INTRODUCTION

The Department of Applied Behavioral Science at the University of Kansas supports the practice of protection for human subjects participating in research. The following information is provided for you to decide whether you wish your child to participate in the present study. You may refuse to sign this form and not allow your child to participate in this study. You should be aware that even if you agree to allow your child to participate, you are free to withdraw at any time. If you do withdraw your child from this study, it will not affect your relationship with this unit, the services it may provide to you, or the University of Kansas.

PURPOSE OF THE STUDY

The purpose of this study is to examine the effects of teaching errors on student learning.

PROCEDURES

Participants will be taught four academic tasks involving geography, biology, and astronomy during the course of the study. The tasks will include identifying (1) countries on an outline map, (2) major rivers, (3) local aquatic insects, and (4) pictures of moons orbiting planets in the solar system. Specifically, students will be asked to color in outlines of countries, trace lines representing rivers, and choose colored icons from a bin representing various insects or moons. Throughout the study, participants will earn points, which can be exchanged for rewards. The pool of rewards will be determined by an assessment of your child's preferences and will consist of academically appropriate items (e.g., colorful folders, pencils, coloring supplies, stickers, etc.). During a portion of the instructional sessions, the feedback given following responses will vary in accuracy.

A. Baseline – The purpose of the baseline phase is to determine whether your child has already learned the tasks. No feedback will be given following responses but participants will earn points towards a reward item described above in exchange for appropriate academic behavior (e.g., sitting quietly, paying attention).

B. Imperfect Instruction – During this condition, feedback will be provided following responses. In some cases, correct responses will be followed by praise and points towards a reward and incorrect responses will be followed by a non-corrective statement such as “Nice try.” Following a proportion of answers, the feedback provided will be inaccurate. That is, occasionally, praise and points will be given following an incorrect answer and praise will be omitted and replaced by a noncorrective statement following a correct answer. The proportion of trials where an error will be made will vary depending on the task. Each task will be associated with a level of teaching accuracy and errors will be made during 0%, 25%, 50%, or 75% of trials in this condition. The experimenter will not provide the correct answer during any trials. The type of error experienced by participants is dependent on their answers. If a student consistently provides correct answers, the errors experienced will consist of the omission of praise and rewards. If a student consistently provides incorrect answers, the errors experienced will consist of praise and rewards for those incorrect answers. Typically, the errors experienced by participants will be a combination of the two.

C. Perfect Instruction – During this condition, errors will no longer be made. That is, all correct answers will be followed by praise and a point and all incorrect answers will be followed by a

non-corrective, neutral statement. The experimenter will not provide the correct answer following incorrect responses.

Previous research has shown that errors in instruction similar to the procedures above will result in a delay of learning for the material presented under imperfect instruction conditions. The sequence of conditions in this study is designed to determine the size of the effect caused by teacher error by measuring the length of the delay to mastery after errors are removed. Each session will be conducted in your child's school and will consist of 30-60 minutes of instruction time. Participants will be given 5-minute breaks every 15 minutes. Thus, one session will last about 45-75 minutes (instruction plus breaks). During breaks, participants will be permitted to engage in preferred leisure activities and no requests or demands will be placed. Preferred leisure activities will be determined by providing options to the student consisting of activities that are typically available to him/her during a regular school day. Imperfect instruction will be ended when a clear, consistent difference is observed between performance on the tasks or when a maximum of 6 hours of exposure to false feedback has elapsed. One session will be conducted per day, 2-3 days per week. The times during which sessions are conducted will be determined by working with teachers to identify times that will least interfere with your child's normal educational activities. Prior to each session, we will ask your child whether they would like to work with us and your child may refuse to participate on any given day or ask to stop at any time during a session. Requests to stop will be honored without penalty. Additionally, if your child chooses not to participate on three consecutive days, he/she will be excluded from further participation in the study. We estimate that the study will take approximately 2 weeks to complete depending on how many sessions can be scheduled per week. We anticipate the study lasting approximately 4-6 sessions, making a total time commitment away from normal educational activities no more than 8 hours (including breaks). All sessions will be video recorded. These recordings will be used only by the research team and will not be displayed or released to others without your explicit written consent.

ALTERNATIVES TO PARTICIPATION

You may choose not to allow your child to participate in this study.

RISKS

There are no physical risks associated with participation in this study. Other risks are minimal and may include shyness about working in the presence of the researcher or frustration due to the inconsistency of the feedback during a portion of the study. Your child may refuse to participate during any given day or ask to stop at any time during a session. If your child does not want to participate on three consecutive days, he/she will be excluded from further participation in the study. Additionally, there is a slight risk that your child will learn incorrect information as a result of the inaccurate feedback delivered during the second phase of the study. However, during the final phase of the study, only accurate feedback will be given and the perfect instruction condition will continue until your child has learned and mastered the correct information. Similar research has demonstrated that when errors are made during teaching, the student's learning of that material was delayed by about 4.5 hours of instruction. Should your child choose to terminate their participation prior to mastering the correct information, you will be given a copy of the materials indicating the correct answers. You may provide your child with the correct information if you are concerned that they have retained false information.

BENEFITS

Direct benefits to your child will be minimal, including being able to perform the tasks taught during the study. Benefits to society will be greater and include a better understanding of how errors during instruction influence student learning and whether a longer-term impact occurs as a result of previous exposure to less than perfect instruction. The results of this study will contribute to the literature on treatment integrity and procedural fidelity.

PAYMENT TO PARTICIPANTS

No payment will be made to participating children or caregivers.

PARTICIPANT CONFIDENTIALITY

Participant and caregiver privacy will be protected. All research materials and records will be kept confidential. Paper materials will be kept in a locked cabinet within a locked office to which only the research team will have access. Digital materials, including video recordings, will be kept on a password-protected computer or on a secure, encrypted server. No identifying information will be disclosed in any publication or presentation resulting from the study. Rather, participants will be assigned code names or aliases. At no time will your identity or the identity of your child be released or made available without your explicit, written permission or where required by law.

Permission granted on this date to use and disclose your information, excluding names, remains in effect indefinitely. By signing this form, you give permission for the use and disclosure of your child's information, excluding name, for the purposes of this study at any time in the future.

REFUSAL TO SIGN CONSENT AND AUTHORIZATION

You are not required to sign this Consent and Authorization form and you may refuse to do so without affecting your right to any services you are receiving or may receive from the University of Kansas or to participate in any programs or events of the University of Kansas. However, if you refuse to sign, your child cannot participate in this study.

CANCELLING THIS CONSENT AND AUTHORIZATION

You may withdraw your consent to allow participation of your child in this study at any time. You also have the right to cancel your permission to use and disclose further information collected about your child, in writing, at any time, by sending your written request to: Jason M. Hirst (see address below). If you cancel permission to use your child's information, the researchers will stop collecting additional information about your child. However, the research team may use and disclose information that was gathered before they received your cancellation, as described above.

QUESTIONS ABOUT PARTICIPATION

Questions about procedures should be directed to the researcher(s) listed at the end of this consent form.

PARTICIPANT CERTIFICATION:

I have read this Consent and Authorization form. I have had the opportunity to ask, and I have received answers to, any questions I had regarding the study. I understand that if I have any

Appendix G

Assent Procedure

Assent Script

Prior to beginning each session, the investigator will obtain verbal assent from the participant. The following script will be read:

“We’re going to [continue to] learn about some interesting things like geography and nature. You can earn some prizes like *names of items available* if you work with me. If you would like to stop at any time, let me know. Would you like to work with me for a while?”

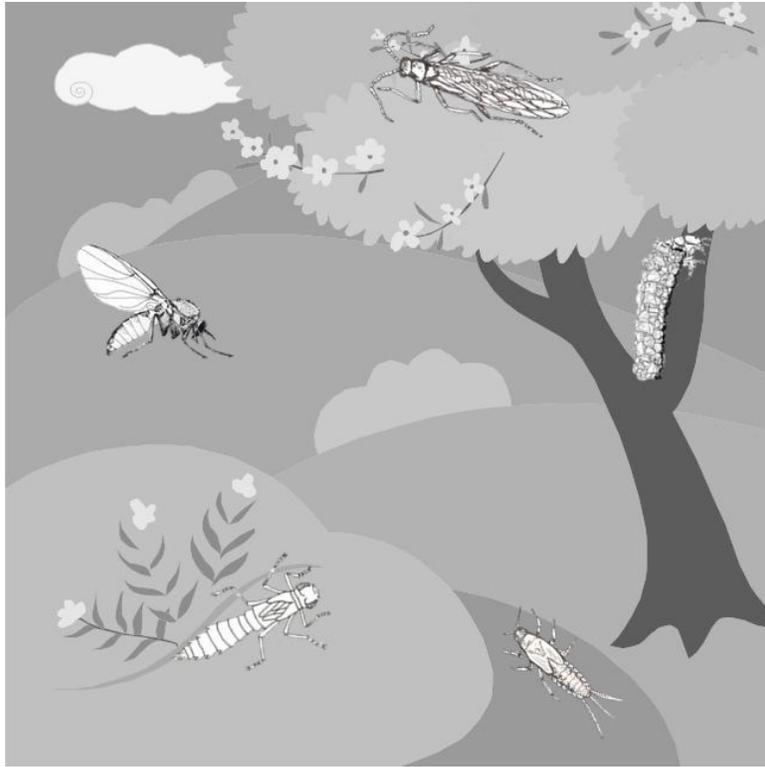
Termination of Session

If at any time the participant verbally expresses a desire to stop working, the session will be terminated for that day. Sessions will also be terminated if the participant engages in sustained problem behavior such as crying or tantrums for 30 seconds during a session.

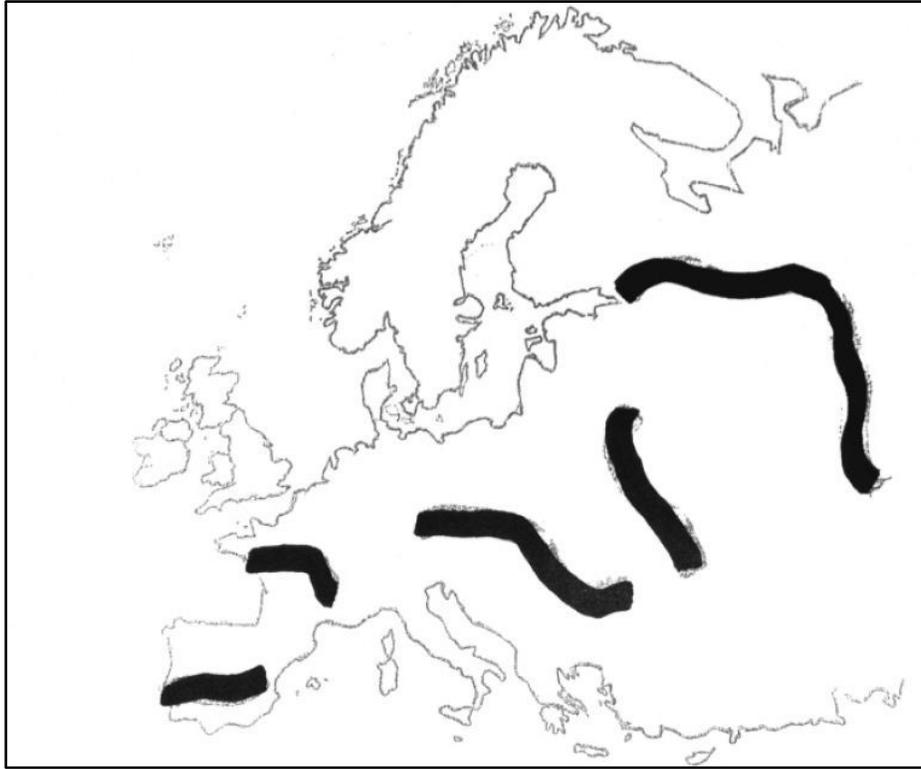
Removal from Study

If three consecutive sessions are terminated, the participant will be removed from further participation in the study.

Appendix H



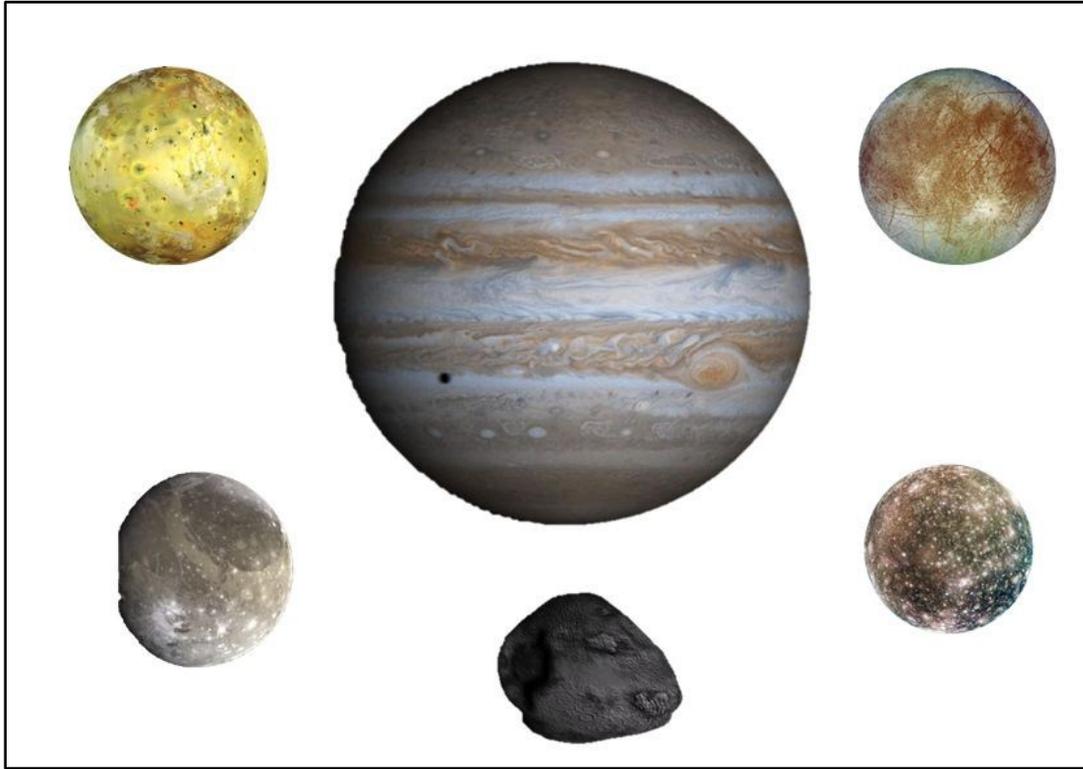
Appendix I



Appendix J



Appendix K



Appendix L

ID	Date	Condition

Condition	Trial	Target	caddisfly	danube	turkey	io
		FB Type	25%	50%	75%	100%
BL	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
IF	17					
	18					
	19					
	20					
	21					
	22					
	23					
	24					
	25					
	26					
	27					
	28					
	29					
	30					
	31					
	32					

Vita

NAME OF AUTHOR: Jason M. Hirst

EMAIL: jmhirst@ku.edu

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED

Pennsylvania State University, University Park, PA
University of Kansas, Lawrence, KS

DEGREES AWARDED:

Bachelor of Science in Psychology, 2008, Pennsylvania State University

AWARDS AND HONORS

First Place, Graduate Research Competition, Sigma Xi, University of Kansas Chapter

PROFESSIONAL EXPERIENCE

Instructor, Department of Applied Behavioral Science, University of Kansas, 2012 - 2013

Graduate Teaching Assistant, Department of Applied Behavioral Science, University of Kansas, 2011 - 2013

Practicum Student, Community Living Opportunities, Children's Network, Lawrence, KS, 2011

Student Hourly, Center for Teaching Excellence, University of Kansas, 2011