12-2022

An Investigation of Lab-Based Research Procedural Fidelity: The Relationship between Experimenter Infant-Directed Speech, Temperament and Language Proficiency

Tess A. Simpson
East Tennessee State University

An Investigation of Lab-Based Research Procedural Fidelity: The Relationship between Experimenter Infant-Directed Speech, Temperament and Language Proficiency

Tess A. Simpson

An Investigation of Lab-Based Research Procedural Fidelity: The Relationship between Experimenter Infant-Directed Speech, Temperament and Language Proficiency

A thesis

presented to

the faculty of the Department of Psychology

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Arts in Psychology

by

Tess Simpson

December 2022

Wallace E. Dixon, Jr., Ph.D., Chair

Alyson Chroust, Ph.D.

Eric Sellers, Ph.D.

Keywords: Fidelity, temperament, language proficiency

ABSTRACT

An Investigation of Lab-Based Research Procedural Fidelity: The Relationship between Experimenter Infant-Directed Speech, Temperament and Language Proficiency

by

Tess Simpson

The purpose of the present study was to investigate whether developmental researchers were influenced in the laboratory by the characteristics of children who participate in their research. I hypothesized that experimenters, as social partners, would adapt their speaking and other behaviors to the child's perceived temperamental profile and language proficiency. I specifically focused on whether experimenters would adhere to the experimental laboratory procedure of two elicited imitation tasks, Feed Bear and Make a Rattle, in an archival dataset. Participants included 61 primarily white 15-month-olds. Coders transcribed infant directed speech (IDS) and analyzed transcriptions for total words, words per sentence, and percentage of words with six or more letters. The present study revealed differential correlational findings across temperamental dimensions, experimenter IDS, and elicited imitation tasks. An investigation of this kind provides new information concerning procedural fidelity and how experimenters may be influenced by their child research participants.

ACKNOWLEDGMENTS

Thank you to my advisor and mentor, Dr. Wallace Dixon, Jr., for his support as I undertook this intellectual endeavor. I cannot say I completed this project without his conceptual and scholarly guidance. Thank you for pushing me to be a better writer and researcher.

I want to also thank my committee members, Dr. Alyson Chroust and Dr. Eric Sellers for their commitment to my success through their invaluable feedback and support. Additionally, thank you to the members of The Program for the Study of Infancy that supported my coding efforts. I could not have taken on such a task without your assistance. Your tenacity and excitement for learning encourages me to continue.

Lastly, thank you to my family, my partner, and my friends for their never-ending love and support. Without them, I would not have completed this project with the same zeal or zest. Specifically, to my parents, Dawn and Mark, thank you for holding me up into the light my whole life and for giving me the love to break out and make a place for myself. To Lily, thank you for making me laugh and bringing me into the present moment. To Eric, thank you for your inexhaustible love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1. Introduction

In human-subjects lab-based research, procedural fidelity can be compromised by the extent to which researchers are unconsciously or inappropriately influenced by research participants. The purpose of the present study was to investigate whether developmental researchers in particular are influenced in the laboratory by the children who participate in their research. Specifically, I attempted to extend the bidirectional fidelity findings of Dixon et al. (2021) to a new archival data set. In the following sections, I review the extant literature which supports the possibility that experimenter fidelity may be compromised by individual characteristics of the child. I begin by reviewing some of the relevant literature pertaining to research involving adult participants, followed by the limited research involving child participants. I conclude with a study that allows for a direct test of this possibility using an archival data set involving toddlers.

Throughout developmental science, it appears that lab-based researchers typically assume that experimenters are not differentially influenced by the unique characteristics of their research participants. At least it is extremely uncommon for these researchers to publish fidelity data that demonstrate procedural validity. A review of the literature using search terms including "fidelity," "experimenter bias," "experimenter-child interaction," and "child effects," returned only a handful of studies that directly bore on experimental procedural fidelity in developmental research (i.e., Dusek, 1972; Jones & Cooper, 1971; Page, 1971; Smith & Whitney, 1987).

On the other hand, experimental fidelity has been the subject of a great deal of research in the adult literature. Obviously, the behavior of the experimenter impacts the behavior of research participants, or else experiments involving experimenters would be ineffectual in human subjects research. However, it is probably equally important to investigate the extent to which the behaviors of the experimenter, especially those that are influenced by the research participants, impact research outcomes in non-fidelitous ways. Jones and Cooper (1971), for example, found that adult participants who received more eye contact from experimenters reported more happiness. This kind of effect might compromise experimental fidelity to the extent that a study employs introverted versus extroverted experimenters, for example. Similarly, Page (1971) found that experimenter smiles and glances influenced how likely participants were to use "I" and "We" in a sentence-building task. And there is a huge literature in which experimenter expectancy effects (i.e., when participants' responses are shaped by experimenter expectations) have been shown to influence outcomes in human subjects research (Rosenthal & Rubin, 1978).

Experimental fidelity can also be impacted by the bidirectional interactions between experimenters and participants. In one study, although the bidirectional effects on experimental fidelity were not explicitly addressed, Chartrand and Bargh (1999) found that non-conscious mimicry of the postures, mannerisms, and facial expressions of social partners facilitated smoother interactions and increased liking between a confederate and participant. Additionally, Chartrand and Bargh reported that dispositionally empathic participants mimicked their social partners to a greater extent than those who were less empathic. This kind of socially bidirectional effect can compromise experimental fidelity to the extent that the study has both empathic participants and experimenters, while also relying on experimenter and participant interactions to perform tasks, for example. One could make an argument, then, that without controlling for participant and experimenter empathy, the actions of the participant may not accurately reflect the participant's response to an experimental task, thus potentially compromising the fidelity of the study.

Although few studies specifically investigated child effects on experimenter performance, it seems clear that uncontrolled experimenter effects can impact developmental research outcomes. To illustrate, when attempting to replicate an experiment on play experience, Smith and Whitney (1987) argued that a lack of controls for experimenter effects was responsible for previous significant results in a sample of 4-year-olds. Specifically, they were referring to several past studies that showed a positive link between play experience and divergent, creative thinking in children (Dansky, 1980; Dansky & Silverman, 1973; Dansky & Silverman, 1975; Li, 1978). Whitney and Smith attempted to replicate these results while incorporating blind procedures. In their study, trained experimenters were blind to study hypotheses and treatment groups. Smith and Whitney found that the earlier studies' significant results were not replicable under these conditions and proposed that the previous significant finding of an association between play history and associative fluency was due to experimenter effects.

Similarly, Dusek (1972), using a sample of 9- to 10-year-olds, aimed to illustrate the presence of experimenter-bias effects using a motor task in test-anxious boys and girls. Dusek found that when experimenters were primed to bias girls' abilities over boy's abilities in a marble dropping task, low test-anxious children were influenced by the experimenters' expectations. Specifically, Dusek found that the low-anxious boys decreased their rates of response whereas the low-anxious girls increased their rates of response when the experimenters were biased towards the girls. These two studies showcase how experimenter bias and internal beliefs can influence unintentional violations of experimental fidelity as well as the outcomes of the study.

Despite the paucity of research on child effects on experimenter behaviors, there is considerable literature on the effects of children on their social partners more generally, especially in the context of bidirectional influences in family settings (Pardini, 2008). For example, Vallotton (2009) found that infants influenced their caregivers' responsiveness and quality of care through their own personal characteristics (i.e., gender, age, and communicative behaviors). She found that when an infant used clear communicative behaviors to respond specifically to their caregiver's communication attempts, the caregiver was more responsive overall. Similarly, Snell and colleagues (2015) found that children played a role in shaping their own experiences through impacting their caregivers' behaviors and expectations. Specifically, they reported that toddlers who were perceived by their non-parental caregivers to have lower cognitive ability received less language stimulation when compared to children perceived to have typical cognitive ability. Such bidirectional interactions in the context of social relations can allow us to infer the potential for similar bidirectional effects in the laboratory, and further, to infer the possibility of child effects on experimenter fidelity in laboratory settings.

Language learning is powered by the shared social relationship between an infant and the caregiver, especially in the context of serve and return interactions wherein the infant and caregiver are mutually responsive to one another's actions and verbalizations (Golinkoff et al., 2015; Landry et al., 2002). Of course, any given dyadic interaction may be directed in varying ways by either social partner; and social partners can differ considerably from one another. For example, despite the infant's participation in both social exchanges, Kwon and colleagues (2013) found that mothers generally talked more to their children than did fathers, while fathers showed denser language usage than did mothers. When fathers spoke, their language tended to have a higher ratio of unique words to the total number of words than did the mothers.

Children's linguistic contributions to an interaction also vary, sometimes as a result of the local environment. Bornstein and colleagues (2000) investigated the relationship between a child's language usage in the lab and the home. They found that children's total utterances and mean length of utterances were similar in the laboratory and the home but differed as a function of their familiarity with their social partner. The children spoke less frequently and in less differentiated ways towards the researcher than towards their mothers. If familiarity with a social partner influences how comfortable children are when talking with experimenters, it stands to reason those experimenters who have had longer or more frequent interactions with a child will be more conversant with that child than with a less familiar child. Under such circumstances, different children will receive different linguistic exposure merely as an artifact of the child's familiarity with the experimenter, rather than as a consequence of the experimental procedure. Given the research that showcases how the behaviors of laboratory experimenters are affected by adult participants (i.e., Chartrand & Bargh, 1999; Jones & Cooper, 1971), coupled with the bidirectionality findings from research involving caregivers and children (Pardini, 2008; Snell et al., 2015; Vallotton, 2009), it stands to reason that in developmental studies involving laboratory procedures, the fidelity of experimenter procedure-following could very well be impacted by infant behavior.

Two infant characteristics which are likely to impact experimenter fidelity include temperament and linguistic proficiency. I first address temperament and then turn to linguistic proficiency. Rothbart and Derryberry (1981) define temperament as comprising individual differences in children's reactivity and self-regulation. In Rothbart's framework (e.g., Rothbart & Bates, 2006), temperament encompasses differences in how infants react to the environment such as through their duration, latency, and intensity of attentional and emotional responsiveness to environmental stimuli. Additionally, Rothbart's temperament framework includes how infants regulate themselves behaviorally, attentionally, and emotionally in the environment.

As a developmental construct, children's temperaments are presumed to influence the quality of their environmental and social experiences (Ganiban et al., 2011; Kiff et al., 2011; Sanson et al., 2004). Among other things, children's temperaments are assumed to affect how children's social partners interact with them. Children with easier temperaments, defined here as having high effortful control and low negative affectivity, should be easier to engage with than children with more difficult temperaments, defined as having high negative affectivity and low effortful control. Indeed, Thomas and colleagues (1982) found a significant correlation between a child's difficult temperament and maternal disapproval and rejection, which, although not proof of a causal link, is consistent with the possibility of a causal link.

These differential impacts of temperament on children's social partners should manifest across multiple developmental domains. Research on the impact of children's temperament on their own development has begun to accrue across a few of these domains. For example, purported temperamental influences have been reported within research on children's joint attention with both caregivers and strangers (Salley & Dixon, 2007; Todd & Dixon, 2010; Vaughan et al., 2003). Higher levels of temperamental negative affectivity, or the tendency to experience and display negative emotions, was found to correlate with lower levels of joint attentional bids from social partners in infancy (Morales et al., 2000; Vaughn et al., 2003) and in toddlerhood (Salley & Dixon, 2007). Joint attention refers to the shared focus of two individuals and serves as a social referencing skill.

In one of the few laboratory-based fidelity studies, Dixon, Driggers-Jones, and Robertson (2021) found that infant temperament was associated with experimenters' duration of looking time in a six-trial gaze-following procedure. In the study, experimenters were trained to alternate 8-second looks to target objects on either the right or the left (depending on the trial number).

Despite being blind to infant temperament status, the experimenters looked significantly longer with infants whose caregivers rated them as high on effortful control or surgency. Dixon and colleagues suggested that infant temperament could have influenced the experimenters' looking behaviors.

Linguistic proficiency may also influence experimenters' behaviors. Linguistic proficiency, or the ability to comprehend and express language, may affect how caregivers socially engage with the child. Snell et al. (2015) reported that children with higher language proficiency at 15 months received more non-parental caregiver input at 24 months. This relationship was also evident between 24 and 36 months. Additionally, children who received more non-parental caregiver input at 15 months also demonstrated more language proficiency at 24 and 36 months, suggesting a longitudinal bidirectional effect. The findings of Snell et al. illustrate the transactional relationship between nonparental caregivers and children, at least with respect to linguistic proficiency.

In sum, it makes sense to hypothesize that experimenter fidelity to protocols in the laboratory setting can be influenced by infant characteristics like temperament and linguistic proficiency. Although there are a variety of experimenters' behaviors that can be impacted, such as the duration of experimenters' standardized actions as in Dixon et al. (2021), one particularly interesting behavior that may be impacted is experimenters' speech to infants. Speech is the primary means through which experimenters communicate with babies about laboratory tasks and expectations. Thus, if infant-directed-speech (IDS) varies as a function of the characteristics of the baby, then infant research participants are not all being treated equally by the experimenters, and by many definitions, experimenters are not adhering fidelitously to laboratory protocols.

With respect to temperament, it is unclear exactly how such infant characteristics may impact experimenter IDS. Based on temperament theory (e.g., Ganiban, et al., 2011; Kiff et al., 2011; Rothbart & Bates, 2006, Rothbart & Derryberry, 1981), it stands to reason that temperamentally easy children might facilitate experimenter engagement, thus resulting in a heightened sophistication of experimenter IDS coupled with a reduced frequency in the need for it. Cooperative babies might just need less linguistic guidance generally. In contrast, children with difficult temperaments might reduce the ease of engagement, causing experimenters to speak more frequently and in less complex ways.

Regarding infant linguistic proficiency, it makes sense that experimenters would respond linguistically to relatively precocious babies by using more complex language than they would use with less linguistically proficient infants. As already discussed, it is known that children with higher language skills have been found to elicit more caregiver input (Snell et al., 2015). This evocative effect could increase the talkativeness and sophisticatedness of experimenter engagement. In contrast, less linguistically precocious babies might receive less complex and frequent engagement.

In the present study, I hypothesized that experimenters' adherence to experimental protocols, particularly with respect to their IDS, will vary as a function of children's temperament and linguistic competence. Specifically, I anticipated that two aspects of experimenter IDS would be influenced: how *much* experimenters talked to the children and how *sophisticatedly* they talked to the children. The former may be measured by a simple count of how many words the experimenter used when interacting with the children. The latter may be indexed by measures of linguistic complexity, including sentence density (i.e., the number of words per sentence) and the use of large words (i.e., the percentage of words with six or more letters).

In terms of operational definitions, temperament can be defined by the various dimensions included in the Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006). One could define temperamental easiness as comprising high scores on the superdimension of effortful control or either of two of the subdimensions of surgency, namely sociability and positive anticipation. Similarly, temperamental difficulty could be defined as comprising high scores on the superdimension of negative affectivity or either of two of the subdimensions of surgency, namely impulsivity and activity level. Either of the latter could be described as challenging to manage, at least from the point of view of the primary caregiver. To be sure, subdimensions of surgency are so called because they empirically "hang together." However, they can also be conceptualized as aligning differentially with easiness versus difficultness. The dimensions reflecting positive affectivity (i.e., sociability and positive anticipation) might arguably better align with temperamental easiness, whereas the dimensions reflecting lack of behavioral control (i.e., impulsivity and activity level) might arguably align better with temperamental difficultness. One can conceptualize the fifth subdimension of surgency, high intensity pleasure, as both a facet of easiness and difficulty. On one hand, infants who have high levels of high intensity pleasure may be temperamentally easier due to their excited positive affectivity. On the other hand, infants high in high intensity pleasure may be considered more difficult due to frequent outward bursts of excitation. Because of the potentially differential alignments of high intensity pleasure, I treated this subdimension as an exploratory measure.

In terms of children's linguistic proficiency, a common measure employed in the extant literature is the MacArthur-Bates Communicative Development Inventory (MBCDI; Fenson et al., 2007). The MBCDI is a parent-reported checklist that comprises measures of vocabulary across the lexical spectrum, including nouns, predicates, and "closed class" words (e.g., pronouns, prepositions, and conjunctions). For present exploratory purposes, a simple measure of total productive vocabulary suffices to index children's linguistic precocity in very early childhood.

In sum, I hypothesized that experimenters' IDS, as reflected in total word use, words per sentences, and percentage of words with six or more letters, to be influenced by children's temperamental easiness, temperamental difficulty, and linguistic proficiency. As well, in an effort to replicate Dixon et al. (2021), I explored the effects of infant temperament on a nonlinguistic experimenter behavior, namely the length of time experimenters exposed infants to novel stimuli during a stimulus familiarization episode. Based on Dixon et al., there is reason to believe that experimenters' time estimates during laboratory protocol administration may be influenced by infant characteristics. But while Dixon et al. employed experimenter visual gaze duration as their measure of interest, in the present investigation I employed a more complex experimenter behavior involving the presentation, display, and removal of novel stimuli. For this behavioral measure, I hypothesized that children with easier temperaments may be content exploring the objects and experimenters would be inclined to let them continue exploring beyond the allotted time. However, with children with more difficult temperaments, experimenters may move through the task at a quicker pace to limit difficult behavior, thus resulting in less familiarization time. These hypotheses are aligned with Dixon et al., who found that experimenters looked significantly longer with children rated high on effortful control. In sum, I hypothesized that the temperament dimensions of easiness (effortful control, sociability, and positive anticipation) would correlate positively with experimenter duration of familiarization time. I expected the inverse correlation between duration of familiarization and dimensions of temperamental difficulty (negative affectivity, activity, and impulsivity).

Chapter 2. Method

Participants

Data used in the present study were procured from an archival data set housed in the Program for the Study of Infancy at East Tennessee State University. In the data set, all participants were recruited from rural Southern Appalachia. Sixty-six primarily white, middle-class toddlers (M = 15.52 months, SD = 0.47 months) and their parents (61 mothers, 5 fathers) were recruited through local newspaper birth announcements. Approximately half of these participants returned 6 months later for a second lab visit (N = 32, M = 21.77 months, SD = 0.67), although these longitudinal data are not relevant for present purposes.

Five participant videos were either (1) not recorded or (2) experimenters did not perform the task, resulting in 61 usable participant videos. These participants were comprised of 34 male and 27 female, primarily white toddlers (M = 14.86 months, SD = 2.06).

Materials, Measures, & Procedure

Infants' Temperament

The Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006) was used to assess temperament. The ECBQ measures temperament in 18- to 36-month-olds using parent-report and utilizes 201 items on an 8-point Likert-type scale (1 = never, 2 = very rarely, 3 = less than half the time, 4 = about half the time, 5 = more than half the time, 6 = almost always, 7 = always, 8 = not applicable). Although the toddlers in the present investigation were younger than the targeted age range of the ECBQ, Putnam (personal communication, June 15, 2015) recommends the ECBQ for this age group. In completing the ECBQ, caregivers rate the frequency of specific child behaviors over the previous two weeks. The 201 items are then subsumed into 18 subdimensions including, impulsivity, activity level, high intensity pleasure, sociability, positive anticipation, discomfort, fear, motor activation, sadness, perceptual sensitivity, shyness, soothability, frustration, inhibitory control, attention shifting, attention focusing, low intensity pleasure, and perceptual sensitivity. As shown in Table 1, the subdimensions are further subsumed into three superdimensions: negative affectivity, surgency, and effortful control (Putnam et al., 2006). I treated easy and difficult temperaments at the dimensional level. I considered effortful control and two subdimensions of surgency (sociability and positive anticipation) as indexing temperamental easiness. I considered temperamental difficulty to be made up of negative affectivity and two additional subdimensions of surgency (impulsivity and activity level). As stated above, I treated the last subdimension of surgency, high intensity pleasure, as an exploratory dimension.

Table 1

Superdimensions of EBCQ

Negative Affectivity	*Effortful Control*	*Surgency*
Discomfort	Inhibitory Control	Impulsivity
Fear	Attention Shifting	Activity Level
Motor Activation	Attention Focusing	High Intensity Pleasure
Sadness	Cuddliness	Sociability
Perceptual Sensitivity	Low Intensity Pleasure	Positive Anticipation
Shyness		
Soothability *		
Frustration		

*Reverse scored

Children's Linguistic Proficiency

The MacArthur-Bates Communicative Development Inventory: Words and Gestures (MBCDI-WG; Fenson et al., 2007) was used to assess linguistic proficiency. Prior to arriving at the lab, parents were sent the MBCDI-WG to complete in their own homes. In the MBCDI-WG, parents documented whether children "understand" or "understand and say" 434 vocabulary items. From this instrument, I used total productive vocabulary as an indicator of toddler linguistic proficiency.

Experimenter Language Talkativeness and Sophisticatedness

Toddlers and their parents visited the laboratory and were audio/video recorded engaging with experimenters in various age-appropriate tasks during sessions that lasted approximately 60 minutes. Ten different experimenters lead the sessions across the participant sample. The number of participants completed by each of the experimenters can be seen in Table 6.

In the present study, I investigated experimenter talkativeness (total words) and sophisticatedness (words per sentence and percentage of words with six or more letters) during a "Feed Bear" and a "Make a Rattle" task. In the Feed Bear task, the experimenter presented children with a play seat, a teddy bear, a napkin, and a spoon, allowing the child to explore the objects for 60 seconds. After the familiarization period, the experimenter placed the teddy bear in the seat, then pretended to feed the bear with a spoon, and lastly, wiped the bear's mouth with the napkin. The experimenter then gave the objects to the child and prompted them to feed the bear too.

During this task each experimenter spoke to each child before, during, and after the modeled demonstration. The before-task language was deemed the "prologue phase," language during the task demonstration was called the verbal “narrative phase,” and the post-task language, while retrieving the objects, was called the “epilogue phase.” The verbal narrative script employed by the experimenters went thusly: "Watch what I can do. I am going to feed the bear. The bear is so hungry. I am going to put him in the chair. Now I am going to give him some cereal. Here bear, here is some cereal. It's so good, isn't it? Yummy. All done." Experimenters repeated the narrative script and accompanying demonstration twice before continuing the task. Examples of language in the prologue phase were “look at that” or “look at what I’ve got.” Examples of language in the epilogue phase were “good job” or “want some more toys?”

In the Make a Rattle task, the experimenter presented two nesting cups and a small wooden block to the child, allowing the child to become familiarized with the objects for 60 seconds. After the familiarization period, the experimenter placed the wooden block into one of the nesting cups, placed the other nesting cup over it, and shook it to make a rattle. The experimenter then handed the objects back and prompted the child to make a rattle with the objects.

As with Feed Bear, experimenter's speech to children during the Make a Rattle task consisted of a prologue phase, a verbal narrative phase, and an epilogue phase. The narrative script employed by the experimenters went thusly: “Watch what I can do. I am going to make a rattle. I am going to put this in here. I’m going to cover it up with this cup. Now I am going to shake it. Listen! I made a rattle.” Experimenters repeated this narrative script and accompanying actions a second time before continuing the task. Prologue and epilogue IDS were much the same as in Feed Bear.

To extract experimenter IDS talkativeness and sophisticatedness from these videos, each video was subjected to the following two-pass procedure. First, four independent coders (two coders per task) transcribed the experimenter's IDS during the Feed Bear and Make Rattle tasks. Raters were trained to code IDS on a similar archival data set that contained similar tasks (i.e., a Make Breakfast task and the Make a Rattle task) to reach 80% reliability on 10% of the sample (i.e., 6 infants). Reliability at the word level was defined as the number of words agreed upon out of the total possible words transcribed. Sentence reliability was defined as the number of sentences agreed upon out of the total possible number of sentences. Raters only transcribed experimenter IDS. After language coding was complete, raters completed reliability checks in which they transcribed 10 percent of their partner's videos to ensure they retained reliability. Post-coding reliability remained above 89.93% at the word level and above 85.19% at the sentence level except for the two coders responsible for the Make a Rattle task. The two coders for Make a Rattle reached reliability at the word level but failed to meet reliability on the sentence level. To expedite coding during Make a Rattle, the two coders worked side-by-side to reach sentence agreement by coding all the videos together. They reviewed both transcripts while watching the videos, evaluated disagreements, and discussed until arriving at agreement. If they were unable to agree, I, as the third coder, made the final decision.

The transcripts were then analyzed through the Linguistic Inquiry and Word Count 2015 platform (LIWC 2015; Pennebaker et al., 2015). LIWC 2015 utilizes a dictionary of approximately 6,400 words and word stems to generate an output that reflects the talkativeness (total words) and sophisticatedness (words per sentence and percentage of words with six or more letters) of a transcript. These three measures, total words, words per sentence, and percentage of words with six or more letters, served as dependent variables of IDS for the present investigation.

Duration of Familiarization

In an effort to replicate Dixon et al. (2021), a nonlinguistic measure of experimenter behavior was employed as an additional dependent variable of interest. Adherence to a target duration of object familiarization time (specifically, 60 seconds in each task) was evaluated as a function of infant temperament and linguistic proficiency. To extract this information, a team of two coders recorded familiarization time side by side in both the Make Rattle and Feed Bear tasks. Using this approach was more expedient than training coders to achieve a specific reliability threshold. Coders recorded the overall familiarization time, which was defined as the time from the initial introduction of the task stimuli to their retraction at the completion of the familiarization period.

Chapter 3. Results

I present results both collectively, collapsing across the two tasks, and individually by task. Separating out tasks can help identify potential task peculiarities that would be otherwise obscured by a combined-task analysis.

Descriptive Statistics and Fidelity

Means, standard deviations, minimums, and maximums for duration of task and duration of familiarization time are included in Table 2. Descriptive statistics for total words, words per sentence, and percentage of words with six or more letters are reported in Table 3 and 4 for the tasks collectively, and separated out, respectively. Finally, descriptive statistics for infant temperament and infant total productive vocabulary are reported in Table 5.

Table 2

Total Task Time and Familiarization Time in Seconds

Measures	*M*	*SD*	*Min.*	*Max.*
Average Task Duration	232.79	40.62	139.00	327.00
Feed Bear	239.58	31.85	147.00	303.00
Make a Rattle	225.10	47.45	139.00	327.00
Duration of Familiarization	67.90	18.57	37.00	160.00
Feed Bear	68.26	16.66	45.00	125.00
Make a Rattle	67.53	20.63	37.00	160.00

Table 3

Descriptive Statistics for Total Words, Words per Sentence, and Percentage of Words with Six or More Letters Combined Across Both Tasks

Measures		*M*	*SD*	*Min.*	*Max.*
TW					
	Total	149.15	23.47	106.00	245.00
	Prologue	12.03	8.70	0.00	36.00
	Narrative	115.98	16.10	74.00	185.00
	Epilogue	20.39	12.18	0.00	60.00
WPS					
	Total	5.61	0.64	4.38	8.06
	Prologue	3.45	1.62	0.00	7.00
	Narrative	6.08	0.79	4.77	8.58
	Epilogue	4.84	1.96	0.00	13.00
Six Letter Words					
	Total	1.16	1.08	0.00	4.17
	Prologue	3.75	5.26	0.00	20.00
	Narrative	0.47	0.89	0.00	3.39
	Epilogue	3.38	5.25	0.00	20.00

Note. TW = total words, WPS = words per sentence, Six Letter Words = percentage of words with six or more letters

Table 4

Total Words, Words per Sentence, and Percentage of Words with Six or More Letters by Feed Bear and Make a Rattle

	M	*SD*	*Min.*	*Max.*
Feed Bear				
TW				
Total	159.81	24.27	116.00	245.00
Prologue	11.45	8.94	0.00	35.00
Narrative	122.42	18.56	74.00	185.00
Epilogue	22.94	11.52	6.00	55.00
WPS				
Total	5.29	0.43	4.38	6.48
Prologue	3.47	1.64	0.00	7.00
Narrative	5.58	0.57	4.77	7.00
Epilogue	5.07	1.49	2.86	8.00
Six Letter Words				
Total	1.05	1.00	0.00	4.17
Prologue	3.80	5.40	0.00	20.00
Narrative	0.51	0.92	0.00	3.39
Epilogue	2.73	4.00	0.00	16.67
Make a Rattle				
TW				
Total	141.23	20.08	106.00	189.00
Prologue	12.63	8.56	0.00	36.00
Narrative	109.33	9.51	82.00	127.00
Epilogue	17.77	12.46	0.00	60.00
WPS				
Total	5.93	0.66	5.03	8.06

Prologue	3.42	1.62	0.00	6.50
Narrative	6.60	0.63	5.41	8.58
Epilogue	4.60	2.34	0.00	13.00
Six Letter Words				
Total	1.26	1.16	0.00	3.88
Prologue	3.70	5.19	0.00	17.65
Narrative	0.43	0.88	0.00	20.00
Epilogue	4.08	6.32	0.00	20.00

Note. TW = total words, WPS = words per sentence, Six Letter Words = percentage of words with six or more letters

Table 5

Descriptive Statistics for Temperament and Total Productive Vocabulary

	M	*SD*	*Min.*	*Max.*
Neg. Affectivity	2.72	.47	1.71	3.66
Effortful Control	4.37	.52	2.71	5.60
Pos. Anticipation	4.37	1.04	1.91	6.50
Sociability	5.38	1.03	2.63	7.00
Activity	4.88	.72	2.80	6.50
Impulsivity	5.10	.71	3.44	6.40
High Intensity Pleasure	4.74	1.05	1.88	6.58
Total Productive Vocab	29.92	33.90	2.00	149.00

To evaluate the temperamental profile of infants in the present sample with respect to those in Putnam et al.’s (2006) normative sample, a series of single-sample *t*-tests were conducted. The present sample differed from the normative sample by exhibiting both lower negative affectivity $t(55) = -5.84$, $p < .001$ and activity level, $t(55) = -28.48$, $p < .001$. In other words, infants in the present sample were less difficult than those in Putnam et al. There were no other differences.

Comparing the productive vocabulary of the infants in the present sample with respect to the normative sample in Fenson et al. (1994) is more complicated. Fenson et al. provided a line graph to mark the median number of words produced across age 8 months to 16 months. To evaluate the total produced vocabulary, I estimated, based on eyeballing the line graph, that 15 ½-month-olds produced approximately between 25 and 50 words. The present study's sample had a mean productive vocabulary of 29.92, which lies within this range.

To explore experimenter fidelity, specifically with respect to duration of familiarization, narrative script total words, narrative script words per sentence, and narrative script percentage of words with six or more letters, I conducted a series of one sample t-tests to compare actual experimenter performance, on average, to anticipated performance based on the experimenter scripts. Duration of familiarization should have been 60 seconds in both Feed Bear and Make a Rattle. Narrative script total words should have been 98 words for Feed Bear and 82 words for Make a Rattle. Average words per sentence should have been 5.44 words for Feed Bear and 5.86 words for Make a Rattle. The percentage of words with six or more letters should have been 0 in both cases. LIWC 2015 rounded to the nearest whole number, resulting in a percentage of 0 even though each script contained at least two words with six letters. If the one sample t-tests were significant, I then conducted one-way ANOVAs to identify which of the experimenters were the source of the difference.

Three experimenters were excluded from these analyses due to only having performed the tasks with one participant. Specifically, for the Feed Bear task I excluded Experimenters 5 and 8 from the t-tests and one-way ANOVAs, and for the Make a Rattle task I excluded Experimenters 5 and 10. In Table 6 the number of participants completed by each experimenter in each task can be found.

Table 6

Experimenter Identity and Number of Participants Completed

Experimenter Identity	*Feed Bear N*	*Make a Rattle N*
Experimenter 1	5	4
Experimenter 2	6	6
Experimenter 3	8	5
Experimenter 4	7	4
Experimenter 5	1	1
Experimenter 6	0	3
Experimenter 7	3	0
Experimenter 8	1	4
Experimenter 9	0	2
Experimenter 10	0	1

For familiarization times in the Feed Bear task, the one-sample t-test revealed that experimenters granted significantly longer durations of familiarization to the infants (M = 68.72, SD = 17.14) than was scripted, $t(28) = 2.74$, $p = .011$. I conducted a one-way ANOVA to investigate potential differences among the experimenters in familiarization time during Feed

Bear; however, the ANOVA did not reveal any. For Make a Rattle, there was not a significant difference between the experimenters' durations of familiarization (*M* = 67.64, *SD* = 21.38) and the scripted 60 second familiarization time, indicating that, as a group, they were fidelitous to the script for this task.

In terms of the total words in the Feed Bear task, experimenters on average used more words during the narrative phase (*M* = 123.90, *SD* = 18.03) than they were supposed to, based on the narrative script, $t(28) = 7.73$, $p < .001$. But again, the one-way ANOVA revealed no significant differences among the experimenters. Experimenters also spoke more words in the Make a Rattle task (*M* = 109.00, *SD* = 9.52) than called for by the narrative script, $t(27) = 15.01$, $p < .001$. In this case, however, the one-way ANOVA revealed that there was a statistically significant difference in total words by experimenter identity, $F(6, 21) = 3.02$, $p = .027$. LSD multiple comparisons indicated that Experimenter 2's narrative phase total words (*M* = 117.67, *SD* = 9.37) were significantly higher than Experimenter 1 (*M* = 107.00, *SD* = 3.46), Experimenter 4 (*M* = 105.00, *SD* = 4.24), and Experimenter 8 (*M* = 97.25, *SD* = 13.05). After removing Experimenter 2 from the analysis, the one sample t-test was still significant. Experimenters still employed more total words (*M* = 106.64, *SD* = 8.26) than called for by the narrative script, $t(21) = 13.99$, $p < .001$. However, a second one-way ANOVA without Experimenter 2 revealed no significant differences among the experimenters.

In terms of words per sentence in the Feed Bear task, experimenters did not deviate significantly from the narrative script. On the other hand, experimenters used more words per sentence in Make a Rattle (*M* = 6.55, *SD* = 0.61) in comparison to the narrative script, $t(27) = 5.92$, $p < .001$. The one-way ANOVA did not reveal significant differences between experimenters in terms of words per sentence.

In terms of the percentage of words with six or more letters, in the Feed Bear task experimenters used a higher percentage than was called for by the narrative script [*M* = .54, *SD* = .94); $t(28) = 3.09$, $p = .004$], although there were no differences among the experimenters. Experimenters also used a higher percentage of words with six or more letters in Make a Rattle (*M* = .43, *SD* = .90) when compared to the narrative script, $t(27) = 2.52$, $p = .018$. But again, there were no differences among the experimenters.

In addition to investigating fidelity for adherence specifically to the narrative script, I also conducted a series of one-way ANOVAs simply to explore whether experimenters significantly differed from one another in terms of total words, words per sentence, and percentage of words with six or more letters during either the prologue and epilogue phases (i.e., when there was no procedural script). Investigating experimenter consistency in the absence of a script may highlight individual differences between experimenters in terms of talkativeness and sophisticatedness. When it is found that certain experimenters speak more or less to infant participants than others, there may be reasons to implement procedures to mitigate such differences. However, no differences among experimenters were found, in either Feed Bear or Make a Rattle, for total words or words per sentence in either the prologue or the epilogue phase. There were also no differences in Feed Bear in the percentage of words with six or more letters in the prologue phase. However, there were differences on this variable during Feed Bear in the epilogue phase, $F(4, 24) = 3.15$, $p = .032$. LSD multiple comparisons indicated that Experimenter 4's epilogue phase percentage of words with six or more letters (*M* = 6.53, *SD* = 5.75) was significantly higher than Experimenter 1's (*M* = 2.13, *SD* = 2.04), Experimenter 3's (*M* = 0.63, *SD* = 1.17), and Experimenter 7's (*M* = 0.00, *SD* = 0.00). After removing Experimenter 2 from the analysis, the one-way ANOVA did not indicate any significant differences between experimenters by their use of words with six or more letters. For Make a Rattle, the one-way ANOVAs did not reveal significant differences for the percentage of words with six or more letters in either the prologue or epilogue phase.

The Effects of Infant Characteristics

Infant Temperament

To investigate my hypotheses about the association between infant temperament and characteristics of experimenter behavior, I conducted a series of Pearson Product-Moment correlations (see Table 7) between the experimenter duration of familiarization, the three experimenter IDS measures in each of the experimental phases of prologue, narrative, and epilogue, and infant temperament. Because of the exploratory nature of these analyses, I set alpha to .10 to reduce the probability of a Type 2 error while maintaining a relatively low risk of a Type 1 error. Recall that the dimensions of temperament included in these analyses were those reflecting easy temperament (effortful control, sociability, and positive anticipation) and those reflecting difficult temperament (negative affectivity, impulsivity, and activity level). Lastly, as described in the introduction, I also conducted an exploratory analysis of the association between experimenter performance and the temperament dimension of high intensity pleasure.

As before, for one set of analyses the IDS measures were collapsed across tasks to provide a general picture of the relationship between the expected predictor and outcome variables. Then I conducted analyses separately for the Feed Bear and Make a Rattle tasks to examine whether one or the other task was more responsible for driving any potential relationships observed in the combined-task analyses. In terms of statistical power, G*Power (Faul et al., 2009) indicated that a two-tailed correlation with a sample size of 60 yields adequate power (1 - β = 0.84) to detect an effect size of 0.30. A medium effect size of 0.30 is consistent with Dixon et al.'s (2021) reported effect size in a similar fidelity study involving child temperament. To make sure the present correlations were not driven by the previously excluded experimenters, I conducted the same correlations without Experimenter 5 and 8 in Feed Bear and without Experimenter 5 and 10 in Make a Rattle. When under visual scrutiny, the correlations do not differ substantially from the present reported correlations.

Table 7

Correlations Among Temperament, Duration of Familiarization, Total Words, Words per Sentence, and Percentage of Words with Six or More Letters

	1	2	3	4	5	6	7
Both Tasks							
Familiarization							
	-	-	-	-	-	-	-
TW							
Total	-	-	-	-	-	.26**	-
Prologue	.29**	-	-	-	-	-	-
Narrative	-	-	-	-	-	-	-
Epilogue	-	-	.32**	-.28**	-	.23*	-
WPS							
Total	-	-	-	-	-	-	-
Prologue	-	-	-	-	-	-	-
Narrative	-	-	-	-.25*	-	-	-
Epilogue	-	-	-	-.26*	-	-	-
Six Letter Words							
Total	-	.23*	-	-	-	.29**	-

Prologue	-	-	-	-	-	-	-
Narrative	-	.24*	-	-	-	.25*	-
Epilogue	-	-	-	-	-	.29**	.27*
Feed Bear							
Familiarization							
	-	-	-	-	-	-	-
TW							
Total	.36*	-	.35*	-	-	.36*	-
Prologue	.47**	-	-	-	.34*	-	.40**
Narrative	-	-	-	-	-	-	-
Epilogue	-	-	.44**	-	.45**	-	-
WPS							
Total	-	-	-	-	-	-	-
Prologue	-	-	-	-	-	-	-
Narrative	-	-	-	-	-	-	-
Epilogue	-	-	-	-	-	-	-
Six Letter Words							
Total	-	-	-	-	-	.36*	-
Prologue	-	-	-.38*	-	-	-	-
Narrative	-	-	-	-	-	.35*	-
Epilogue	-	-	-	-	-	-	-
Make a Rattle							
Familiarization							
	-	-	-	-	-	-	-
TW							
Total	-	-	-	-	-	.33*	-
Prologue	-	-	-	-	-	-	-.37*
Narrative	-	-	-	-	-	-	-
Epilogue	-	-	-	-	-	-	-
WPS							

	1	2	3	4	5	6	7
Total	-	-	-	-	-	-	-
Prologue	-	-	-	-	-	-	-
Narrative	-	-	-	-.42**	-	-	-
Epilogue	-	-	-	-.33*	-	-	-
Six Letter Words							
Total	-	.37*	-	-	-	-	-
Prologue	-	-	-	-	-	-	-
Narrative	-	-	-	-	-	-	-
Epilogue	-	.46**	-	-	-	.38*	.39*

Note. Familiarization = duration of familiarization time; TW = total words; WPS = words per sentence; Six Letter Words = percentage of words with six or more letters. 1 = effortful control; 2 = positive anticipation; 3 = sociability; 4 = negative affectivity; 5 = activity; 6 = impulsivity; 7 = high intensity pleasure. Empty cells denote a nonsignificant value. $p < .10$ *, $p < .05$**, $p < .01$***

Temperament x Duration of Familiarization. I originally hypothesized that temperament dimensions of easiness (effortful control, sociability, and positive anticipation) would correlate positively with experimenter duration of familiarization time. Additionally, I expected a negative correlation between duration of familiarization and dimensions of temperamental difficulty (negative affectivity, activity, and impulsivity). However, there were no significant associations between temperament dimensions and duration of familiarization.

Temperament x Experimenter IDS Total Words. In terms of potential associations between infant temperament and experimenter total words uttered, I hypothesized that the temperament dimensions of easiness would be negatively correlated with experimenter total words. However, results revealed that only sociability was positively correlated with total words, and then only in the Feed Bear task during the epilogue phase. Neither effortful control nor positive anticipation were correlated with total words during any phase of either task.

I expected the inverse relationship with IDS for temperament dimensions of difficulty. Specifically, I expected the temperament dimensions of negative affectivity, impulsivity, and activity level to be positively correlated with experimenter total words. However, results revealed that negative affectivity correlated *negatively* with experimenter total word use, at least in the epilogue phase collapsed across tasks. Consistent with expectations, temperamental impulsivity was positively correlated with experimenter overall total words across both tasks, but only in the epilogue phase. Also consistent with expectations was that activity level was positively correlated with total words, in the epilogue phase, but only in the Feed Bear task.

Temperament x Experimenter IDS Words Per Sentence. I next hypothesized that infants' temperamental easiness (effortful control, sociability, and positive anticipation) would be positively correlated with experimenter words per sentence, and that their temperamental difficulty (negative affectivity, impulsivity, and activity) would be correspondingly negatively correlated. However, results revealed that only one temperament dimension correlated with words per sentence. Specifically, and consistent with expectations, negative affectivity was negatively correlated with experimenter's words per sentence collapsed across both tasks and in both the narrative and epilogue phases. Additionally, negative affectivity was negatively correlated with the experimenter's words per sentence in both the narrative and epilogue phases, but only in the Make a Rattle task.

Temperament x Experimenter IDS Percentage of Words with Six or More Letters. In terms of the potential associations between infant temperament and experimenters' percentage of words with six or more letters, I hypothesized that children's temperamental easiness (effortful control, sociability, and positive anticipation) would be positively correlated with the percentage of words with six or more letters. Similarly, I expected that children's temperamental difficulty (negative affectivity, impulsivity, and activity) would be negatively correlated with experimenter percentage of words with six or more letters. There were generally few significant correlations. However, results indicated that impulsivity was positively correlated with the percentage of words with six or more letters, specifically in the feed bear task. Additionally, impulsivity was positively correlated with the percentage of words with six or more letters in the narrative phase collapsed across tasks and the narrative phase in Feed Bear. Impulsivity was also positively correlated with percentage of words with six or more letters in the epilogue phase across both tasks and in Make a Rattle.

Infant Language Proficiency

In terms of the potential associations between infant language proficiency and experimenter behavior, I expected experimenter IDS total words, words per sentence, and percentage of words with six or more letters to positively correlate with infants' total productive vocabulary. As in the previous section, I conducted a series of correlations between total productive vocabulary and experimenter IDS total words, words per sentence, and percentage of words with six or more letters. I conducted each correlation independently for the prologue, narrative, and epilogue phases. As can be seen in Table 8, several significant correlations between children's language proficiency and experimenter IDS were observed at the $\alpha = .10$ level.

Table 8

Total Productive Vocabulary and Total Words, Words per Sentence, and Percentage of Words with Six or More Letters

		Total Productive Vocabulary
Both Tasks		
TW		
	Total	.29**
	Prologue	.24*
	Narrative	-
	Epilogue	.40***
WPS		
	Total	-
	Prologue	-
	Narrative	.34**
	Epilogue	-
Six Letter		
	Total	.36***
	Prologue	-
	Narrative	-
	Epilogue	-
Feed Bear		
TW		
	Total	-
	Prologue	.43**
	Narrative	-
	Epilogue	-
WPS		
	Total	-

Prologue	-
Narrative	-
Epilogue	-.35*
Six Letter	
Total	-
Prologue	-
Narrative	-
Epilogue	-
Make a Rattle	
TW	
Total	.50***
Prologue	-
Narrative	-
Epilogue	.64***
WPS	
Total	-
Prologue	-
Narrative	.51***
Epilogue	.48**
Six Letter	
Total	.60***
Prologue	-
Narrative	.40**
Epilogue	-

Note. TW = total words; WPS = words per sentence; Six Letter Words = percentage of words with six or more letters. Empty cells denote a nonsignificant value. $p < .10$ *, $p < .05$**, $p < .01$***

Language Proficiency x Experimenter IDS Total Words. Experimenter IDS total words were positively correlated with total productive vocabulary when collapsing across both tasks and all phases. Additionally, experimenter IDS total words in the prologue and epilogue phases positively correlated with total productive vocabulary. For the Feed-Bear-specific correlations, experimenter IDS total words positively correlated with child total productive vocabulary, but only in the prologue phase. In Make a Rattle, experimenter total words positively correlated with total productive vocabulary in the epilogue phase.

Language Proficiency x Experimenter IDS Words Per Sentence. Next, experimenter words per sentence across both tasks was positively correlated with infant total productive vocabulary, but only in the narrative phase. One surprising inconsistency was that words per sentence in the epilogue phase was positively correlated with total productive vocabulary in make a rattle but negatively correlated with language proficiency in Feed Bear.

Temperament x Experimenter IDS Percentage of Words with Six or More Letters. Finally, the percentage of words with six or more letters in experimenter IDS, across both tasks and when collapsing across phases, was positively correlated with child total productive vocabulary. However, Make a Rattle was the only task to reveal a positive correlation between total productive vocabulary and the percentage of experimenter's words with six or more letters and then only in the narrative phase.

Chapter 4. Discussion

If infant characteristics like temperament and linguistic proficiency are presumed to influence the quality of their social experiences, then it makes sense to hypothesize that in developmental research, experimenters, as social partners, would adapt their speaking and other behaviors to the child's individual characteristics (Ganiban et al., 2011; Kiff et al., 2011; Sanson et al., 2004). In previous research, Dixon et al. (2021) supported the notion that experimenters can differ in their procedural fidelity as a function of child temperament. In the present study, I attempted to extend their findings to a new archival data set, involving children of a different age, different experimental behaviors of interest, and employing different laboratory procedures.

I was specifically focused on whether experimenters would fidelitously adhere to an experimental laboratory procedure that employed elicited imitation tasks, and further, whether they would adhere to it differentially as a function of the individual characteristics of the infant participants included in the protocol. I had predicted that experimenters' stimulus familiarization times and IDS would be impacted by both infant temperament and infant language proficiency. In terms of stimulus familiarization times, I expected experimenters to grant longer durations of familiarization to children higher in temperamental easiness and grant shorter durations of familiarization to children higher in temperamental difficultness. In terms of IDS, I expected dimensions of temperamental easiness to correlate negatively with experimenter IDS total words and to correlate positively with experimenter IDS words per sentence and the percentage of words with six or more letters. The converse hypotheses were expected to be true for temperamental dimensions of difficulty.

Focusing on just the question of experimental fidelity to protocol, without regard to the question of potential infant influences on experimenters, the results of the present study are valuable in terms of understanding procedural fidelity when implementing experimental protocols. In the simplest sense, experimenters' utterances deviated from the narrative script in both Feed Bear and Make a Rattle. In Feed Bear, experimenters also provided significantly longer novel stimulus familiarization times. Similarly, despite experimenters using more total words than called for in the narrative script in both Feed Bear and Make a Rattle, experimenters in Make a Rattle also used more words per sentence in comparison to the narrative script. These results showcase how assuming procedural fidelity is not the same as controlling or checking for it. Anecdotally, when transcribing the IDS, I noticed that experimenters commonly repeated segments of the narrative scripts, or attempted to correct behavior in the lab setting, rather than adding unrelated conversation into the narrative phase. However, if experimenters are interacting with children differentially, despite having a script, they could change the infant participant's experience in the lab, thus potentially affecting the study outcomes. Although there were differences between the two tasks in terms of fidelity adherence, identifying the potential sources of any task differences would be speculative at best.

One potential explanation for the task differences could be that Feed bear may be a more familiar task than Make a Rattle. In Feed Bear, infants are likely familiar with the motions of pretending to feed a stuffed animal. They are almost certainly familiar with eating events generally. However, in Make a Rattle, the infants were taught a novel sequence of events in order to create a rattle. The concept of a rattle may be familiar, but the act of making one, or at least making one with the specific objects employed in this study, may be novel for most infants. Conceivably, then, the nature of the task may have encouraged experimenters to deviate from the procedural protocol differently as a consequence. However, identifying the source of these differential behaviors would again be speculative.

Artifacts of the Influence of Temperament

One major category of hypotheses related to this project pertained to whether aspects of infant temperament may influence experimental behavior in the lab. My hypotheses regarding associations between infant temperament and experimenter talkativeness were only partially supported. I had proposed that experimenters would need to talk less often to infants who were temperamentally easy; however, only the sociability dimension of temperamental easiness was associated with experimenter total word use, and it was associated *positively*. This finding may be interpretable in hindsight. Perhaps infants who were more temperamentally sociable solicited more communication from the experimenters, thus increasing the total words spoken. On the other hand, I had also proposed that experimenters would need to talk more often to infants who were temperamentally difficult. In this case, I assumed that infants with higher instances of temperamental difficulty would solicit more linguistic management from experimenters, who would thus provide more words to encourage specific behaviors in the experimental setting. This proposal was partially supported; impulsivity and activity level were both positively associated with IDS total words, but only during the epilogue phase. An unexpected finding was that experimenters actually used fewer total words with children who were high in negative affectivity, which is a classic dimension of temperamental difficulty.

Perhaps not surprisingly, and largely paralleling the findings involving talkativeness, my hypothesis regarding potential positive associations between experimenter sophisticatedness, specifically words per sentence, and dimensions of temperamental easiness were not supported. Among the temperamental easiness class of variables, infants' sociability was not associated with words per sentence despite being associated with experimenter talkativeness. I had also proposed that children who were more temperamentally difficult may illicit less sophisticatedness in words per sentence. I presumed that infants with difficult temperaments may receive less sophisticated speech from experimenters due to needing more frequent communication, thus not giving space for sophisticatedness in speech. However, this was only partly supported. Experimenters used fewer words per sentence with children rated higher in negative affectivity across both tasks. Specifically, in Make a Rattle, experimenters used fewer words per sentence during the narrative and epilogue phases when the children were rated higher in negative affectivity than when they were rated lower.

Lastly, my hypothesis that dimensions of temperamental easiness would positively correlate with experimenter IDS percentage of words with six or more letters was unsupported. Experimenters did not tend to use larger words when speaking to children with easier temperaments. However, although not supporting my general hypothesis of an inverse association between temperamental difficulty and sophisticatedness of IDS, children's impulsivity was associated with experimenters' use of large words. That is to say, experimenters used larger words during the narrative and epilogue phases when speaking to children rated as higher in impulsivity. This finding was surprising. I previously conceptualized that experimenters would potentially have a more challenging time communicating with children who scored higher on dimensions of temperamental difficulty such as impulsivity.

Considering the first three IDS measures together, it is not clear why some of the temperamental difficulty measures were negatively associated with IDS total words, why a different one was negatively associated with words per sentence, and why impulsivity was *positively* correlated with percentage of words with six or more letters. All three IDS measures were expected to "hang together" as general indices of IDS. Yet in this sample, there appears to be some dissociability. It may be that different aspects of experimenter IDS are differentially associated with aspects of difficult temperament, but again this would be mere speculation.

Finally, as mentioned in the introduction, I conducted exploratory analyses with the temperamental dimension of high intensity pleasure and experimenter IDS. High intensity pleasure may be conceptualized as a temperamental dimension of either easiness or difficultness, depending on the context. But as with the other temperament dimensions, correlations between high intensity pleasure and experimenter IDS were inconsistent and contradictory. High intensity pleasure had a positive association with experimenter talkativeness in the prologue phase of the Feed Bear task, whereas in the prologue phase of Make a Rattle, experimenters actually talked less when children were rated higher in high intensity pleasure. In contrast, high intensity pleasure seemed to have a positive relationship with experimenter sophisticatedness during the Make a Rattle epilogue phase, specifically related to the percentage of words with six or more letters, despite having no relationship to experimenters' words per sentence. The incongruent results involving this last temperament dimension are puzzling and may reflect differential associations as a function of task context; however, just how task contexts may underwrite these different associations is no less puzzling.

Artifacts of the Influence of Language Proficiency

Unlike with infant temperament, my hypotheses involving associations between infant language proficiency were considerably more straightforward. Specifically, experimenters were expected to talk more and with more sophisticated language with infants who were more linguistically proficient. These hypotheses were fairly well supported. Experimenter IDS total words in the prologue and epilogue phases, when combined across tasks, was positively predicted by infant language proficiency. In looking at the tasks individually, the positive association was in the prologue phase of Feed Bear whereas it was in the epilogue phase of Make a Rattle. Thus, as with temperament, there may be task-specific and phase-specific effects on the associations between infant linguistic proficiency and how many words experimenters use in the lab.

Infant linguistic proficiency was also positively associated with experimenter language sophisticatedness in both tasks, specifically during the narrative portions of the protocols. That is to say, experimenters employed longer sentences during the narrative phase when infants were higher in linguistic proficiency. However, experimenter words per sentence usage was differentially associated with infant linguistic proficiency across tasks in the epilogue phase. In Make a Rattle, experimenters used more words per sentence when speaking to children with higher linguistic proficiency. Whereas in Feed Bear, experimenters' words per sentence were significantly less for the same children.

In terms of sophisticatedness in word length, experimenters employed larger words across both tasks and all phases when infants were higher in linguistic proficiency. However, this association appears to be driven primarily by the narrative phase of the Make a Rattle task, since the corresponding association in the Feed Bear task failed to achieve statistical significance when considered separately.

Considering all three IDS measures together, the nature of these findings is consistent with the hypotheses that experimenters are likely to speak more or more sophisticatedly to infants with higher language proficiency. What is particularly interesting is that although experimenters' expressed language appeared to vary as function of infant linguistic proficiency, the experimenters were blind to the infants' scores on the MBCDI-WG. These findings support the notion that experimenters may be influenced by the individual characteristics of their infant participants; or in this case, infant participants' linguistic proficiency. Although it is impossible to draw a causal conclusion regarding the impact of infant linguistic proficiency on experimenter IDS in the lab, the findings are consistent with a causal conclusion.

Limitations

There are several limitations to the current study. Of course, when considered at the individual task level, sample sizes were relatively small. With lowered statistical power, small samples are likely to lead to spuriously high associations or increase the likelihood of missing an association with a small effect size. Because this study was based on archival data, the sample size was fixed.

Additionally, the archival nature of the data was not designed to answer the present research questions of interest. Due to the archival data, for example, I could not determine whether the experimenters were consciously aware of children's language proficiency or of any of the children's temperamental profiles. These questions were not meant to be addressed in the original study. I also had to work within the bounds of the already present tasks. The original tasks had two versions, one which was used in this study, but also one that contained a distractor element. I was unable to use children's videos that included the distractor, thus limiting sample size for each task.

Another limitation of note is the large number of experimenters in the present study. The number of experimenters adds considerable error variance to the dependent variables of experimenter duration of familiarization and experimenter IDS, as compared to Dixon et al. (2021) who only used two experimenters for roughly the same number of children. Similarly, all experimenters were white-presenting women. So, it is unclear if the present findings would generalize to a sample of experimenters with more diverse ethnic and gender identities. Lastly, infants in the present sample were primarily white and engaged with primarily white experimenters. This again limits generalizability from the present study.

Future Directions

Future research should continue to unpack the bidirectional influences of experimenters and their infant participants in terms of procedural fidelity and the factors emanating from the participants themselves that may impact it. In the present study, there were several instances where a correlation appeared for one task but not the other or where a positive correlation appeared in one task, with a reversed valence in the other. Differential patterns found across the two tasks may be attributed to potential differences in task familiarity, hence future studies may endeavor to ensure greater structural similarities across multiple tasks. The use of multiple familiar tasks, in comparison to the use of multiple unfamiliar tasks, may allow future researchers to separate out effects associated with task familiarity. Further research should be conducted to address the task differences.

Similarly, experimenter familiarity with the child may prove to be an interesting point of future investigation. Potentially, experimenters with longer or more frequent interactions with a child may be more conversant overall. As seen with Bornstein et al. (2000), children spoke similarly in the lab and in the home but differed as a function of the familiarity with their social partner. So, it stands to reason that experimenters who have more time to acquaint themselves with the child will engage in more verbal communication and potentially be more conversant than with a less familiar child. Familiarity could be influenced by the placement of a given task within the lab protocol (i.e., at the beginning of the session, in the middle of the session, or at the end of the session). The placement of the task within the session may increase or decrease familiarity. Understanding the nature of the experimenter's familiarity or exposure to the child could further understanding of procedural infidelity in developmental research by highlighting the causes of infidelity to the procedure.

Conclusion

In the present study, I investigated the possibility that developmental, lab-based experimenters might be susceptible to procedural infidelity, especially as a function of the individual characteristics of the research participants they engage with. These results provide support for this line of inquiry. The results serve as an important reminder of the importance of checking procedural fidelity when relying on human-driven experimental protocols.

References

Bornstein, M. H., Haynes, O.M., Painter, K.M., & Genevro, J.L. (2000) Child language with mother and with stranger at home and in the laboratory: A methodological study. *Journal of Child Language, 27,* 407-420.

Carson, D.K., Klee, T., Lee, S., Williams, K.C., & Perry, C.K. (1998) Children's language proficiency at ages 2 and 3 as predictors of behavior problems, social and cognitive development at age 3. *Communication Disorders Quarterly, 19*(2), 21-30.

Chartrand, T.L., & Bargh, J.A. (1999) The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Dansky, J. L. (1980). Make-believe: A mediator of the relationship be-tween play and associative fluency. *Child Development, 51*, 576-579.

Dansky, J. L., & Silverman, I. W. (1973). Effects of play on associative fluency in preschool-aged children. *Developmental Psychology, 9,* 38-43.

Dansky, J. L., & Silverman, I. W. (1975). Play: A general facilitator of associative fluency. *Developmental Psychology, 11*, 104.

Dixon, W.E., Drigger-Jones, L.P., & Robertson, C.L. (2021) Infant effects on experimenter behavior. *Infant Behavior and Development, 63,* 1-8.

Dusek, J.B. (1972). Experimenter-bias effects on the simple motor task performance of low and high- test anxious boys and girls. *Psychological Reports, 30,* 107-114.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C.B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, 59*(5), 1-185. doi:10.2307/1166093

Fenson, L., Marchman, V.A., Thal., D.J., Dale, P.S., Reznick, S.J., & Bates., E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual.* Paul H. Brookes Publishing Co.

Ganiban, J.M., Ulbricht, J., Saudino, K.J., Reiss, D., & Neiderhiser, J.M. (2011). Understanding child-based effects on parenting: Temperament as a moderator of genetic and environmental contributions to parenting. *Developmental Psychology, 47*(3), 676-692.

Golinkoff, R.M., Can, D.D., Soderstrom, M. & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech, and its effects on early language acquisition. *Current Directions in Psychological Science, 24*(5), 339-344.

Jones, R. A., & Cooper, J. (1971). Mediation of experimenter effects. *Journal of Personality and Social Psychology, 20*(1), 70–74.

Kiff, C.J., Lengua, L.J., & Zalewski, M. (2011). Nature and nurturing: Parenting in the context of child temperament. *Clinical Child and Family Psychological Review, 14*, 251-301.

Kim, S. & Kochanska, G. (2012). Child temperament moderates effects of parent-child mutuality on self-regulation: A relationship-based path for emotionally negative infants. *Child Development, 83*(4), 1275-1289.

Kwon, K., Bingham, G., Lewsader, J., Jeon, H., & Elicker, J. (2013). Structured task versus free play: The influence of social context on parenting quality, toddlers' engagement with parents and play behaviors, and parent-toddler language use. *Child Youth Care Forum, 42*(3), 207-244.

Landry, S.H., Smith, K.E., & Swank, P.R. (2002) Environmental effects on language development in normal and high-risk child populations. *Seminars in Pediatric Neurology, 9*(3), 192-200.

Li, A. & K. F. (1978). Effects of play on novel responses in kindergarten children. *Alberta Journal of Educational Research, 24,* 31-36.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development, 23*(2), 231-236.

Page, S. (1971). Social interaction and experimenter effects in the verbal conditioning experiment. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 25(6),* 463–475.

Pardini, D.A. (2008). Novel insights into longstanding theories of bidirectional parent-child influences: Introduction to the special section. *Journal of Abnormal Child Psychology, 36,* 627-631.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* University of Texas at Austin.

Putnam, S.P., Garstein, M. A., & Rothbart, M.K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development, 29*(3), 386-401.

Rosenthal, R. & Rubin, D.B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences, 3,* 377-415.

Rothbart, M.K., & Bates, J.E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology*: *Social, emotional, and personality development* (6th ed., Vol.1). Wiley.

Rothbart, M., & Derryberry, D. (1981). Development of individual differences in temperament. In M. Lamb & A. Brown (Eds.), *Advances in Developmental Psychology* (pp. 37–86). Lawrence Erlbaum Associates, Inc.

Salley, B.J., & Dixon, W.E. (2007). Temperament and joint attentional predictors of language development. *Merrill Palmer Quarterly, 53*(1), 131-154.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*, 142-170. http://dx.doi.org/10.1046/j.1467-9507.2004.00261.x

Smith, P.K. & Whitney, S. (1987) Play and associative fluency: Experimenter effects may be responsible for previous positive findings. *Developmental Psychology, 23*(1), 49-53.

Snell, E.K., Hindman, A.H., & Belsky, J. (2015) Child effects and child care: Implications for risk and adjustment. *Development and Psychopathology, 27,* 1059-1076.

Thomas, A., Chess, S., & Korn, S. (1982). The reality of difficult temperament. *Merrill-Palmer Quarterly, 28*(1), 1-20.

Todd, J.T., & Dixon, W.E. (2010) Temperament moderates responsiveness to joint attention in 11-month-old infants. *Infant Behavior and Development, 33,* 297-308.

Vallottom, C.D. (2009) Do infants influence their quality of care? Infants' communicative gestures predict caregivers' responsiveness. *Infant Behavior and Development, 32*, 351-365.

Vaughan, A., Mundy, P., Block, J., Burnette, C., Delgado, C., Gomez, Y., Meyer, J., Neal, A. R., & Pomares, Y. (2003) Child, caregiver, and temperament contribution to infant joint attention. *Infancy, 4*(4), 603-616.

VITA

TESS A. SIMPSON

Education: M.A. Psychology, East Tennessee State University, 2022

B.S. Psychology, Tennessee Technological University, Cookeville, TN, 2017

Public Schools, Cookeville, Tennessee

Professional Experience: Instructor, Introduction to Psychology, East Tennessee State University, Fall 2022

Instructor, Introduction to Psychology, East Tennessee State University, Summer 2022

Instructor, Design and Analysis II Lab, East Tennessee State University, Spring 2022

Instructor, Design and Analysis II Lab, East Tennessee State University, Fall 2021

Summer Undergraduate Research Assistant, Tennessee Technological University, 2019

Social Emotion Lab Assistant, Tennessee Technological University, 2019

Select Presentations: Simpson, T.A., Dixon, W.E., Robertson, C.L., Morelen, D.M., & Waltz, O.M. (2021, October). Early Childhood Temperament as a Predictor of Emotion Regulation Difficulties. Presented at the Society of Research in Child Development 2021 Biennial Meeting.

Robertson, C. L. & Simpson, T.A. (2020, October). Introduction of the ETSU Ballad Health Strong BRAIN Institute. Presented at the River Cities Industrial and Organizational Psychology Conference.

Honors and Awards: College of Graduate Studies Excellence in Teaching Award

Teaching Assistant, East Tennessee State University, 2022

Department of Psychology Excellence in Teaching Award

Teaching Assistant, East Tennessee State University, 2022

Hubert & Shirley Chaffin Scholarship, Tennessee Technological University, 2019

Undergraduate Research and Creativity Award (URECA) Grant Program, Tennessee Technological University, 2019

University Service Scholarship, Tennessee Technological University, 2017