The Influence of Shyness on Language Assessment

Liesl Melnick and Sarah C. Kucker

Department of Psychology, Oklahoma State University, Stillwater, OK

Author Note

Sarah C. Kucker https://orcid.org/0000-0003-2210-3599

The authors declare no conflicts of interest.

Funding Statement: The study was funded through an internal Oklahoma State Wentz Fellowship to the first author. All materials and data for the study can be found on OSF: https://osf.io/hwrma/?view_only=905f836347114826a54e57d6d1c10cf1

Correspondences should be addressed to Sarah C. Kucker, 116 Psychology Building, Stillwater, OK 74075. Email: sarah.kucker@okstate.edu

Abstract

Purpose: One understudied component that can influence children's performance on language assessments is shyness. The goal of this study is to examine how shyness affects a child's performance on language assessments which vary in sociability. We hypothesized that accuracy on language tasks would be driven by shyness such that shy children would perform better on non-sociable tasks compared to sociable tasks.

Methods: The procedures followed a quasi-experimental design. 122 participants, ages 17-to-42 months, varied in their temperament and each underwent a series of language tasks. The order of tasks was randomized and included three different language tasks that varied in the social interaction required: a looking task, pointing task, and production task. Parents reported their child's shyness level via the ECBQ. Data was collected via Zoom.

Results: Shyness was compared with participant's accuracy across the three tests while controlling for age and vocabulary percentile. There were significant differences in children's performance across the tasks, with respect to shyness. Shyer children performed much worse on the production task compared to less shy children ($\beta = -.05$, $p = .033$). For the pointing task, shyness interacted with age to impact performance ($\beta = -.37$, $p = .042$). Shyness had no impact on performance on the looking task ($\beta = .03$, $p = .10$).

Conclusions: As shown by these results, shy children and less shy children respond differentially to three methods of language assessment. It is important for clinicians to acknowledge shyness when choosing an appropriate assessment of children's language. Future direction includes examining in-person effects of shyness on language assessment.

Introduction

Children's temperament drives their responses to the environment around them, including what they interact with and the nature of their interaction. Temperament also influences what and how people learn, and how to assess that learning. At the same time, children's behavior and learning are shaped by the environment. This is especially true in language assessment where the style of tasks and settings may vary widely. By the time children are nearly two years old, they have a robust vocabulary. However, in order to demonstrate they know a word, the child's own temperament, the environment, and task at hand together must support a child's ability to bring that knowledge to bear in a given moment.

The current study probes how temperament influences children's response to a series of language assessment tasks. Specifically, unique types of vocabulary and methodological tasks have different demands, especially in the amount of social interaction they require (i.e. if they require talking, gesturing, and other interactions). This variability in tasks may differentially impact performance, leading to incongruent conclusions about a child's vocabulary. The key question here is to what extent a child's temperament, in combination with the task constraints, ought to be considered to obtain a valid and reliable measure of vocabulary.

Relationship between language and shyness

Shyness, a facet of temperament, can be defined as feelings of distress and tension within social situations, specifically apprehension, awkwardness, or self-consciousness (Buss & Plomin, 1984). Researchers refer to shyness as an "approach–avoidance conflict." That is, shy children want social interaction, but at the same time, they are inhibited by their social fear (Asendorpf, 1990). Unlike broader dimensions of temperament like effortful control or surgency, shyness specifically encompasses a child's approach/avoidance to their surroundings, especially social

interaction. It is also important to note that shyness differs from both social anxiety and introversion, such that social anxiety is an extreme form of shyness that also includes high elements of neuroticism (Hofmann & DiBartolo, 2010) and shyness manifests as a combination of both introversion and neuroticism (Briggs, 1988). Additionally, there are several factors that influence the development of shyness, including environmental factors such as parental interactions, peers, sex differences, and other experiences (Eggum et al., 2009).

Shyness begins to develop at a very young age. In infancy, shyness manifests as anxiety, discomfort, and fearfulness of novel or uncertain scenarios, including strangers (Buss & Plomin, 1984, Putnam et al, 2006), developing to be more about behavioral inhibition in social situations by the time children hit adolescence (Putman & Rothbart, 2006). During their second year of life, children develop a sense of self and an understanding of social behavioral standards, allowing them to experience self-conscious emotions and anxiety over other people's evaluation (Emde et al., 2001).

The same period in which shyness begins to emerge, children's language skills are rapidly developing, so it is not surprising that the two are related. For instance, temperament types such as positive emotionality, attention, and social orientation are related to a larger vocabulary size (Usai et al., 2009 & Dixon & Shore, 1997). Additionally, shyness is related to poorer language skills (Evans 1993, Slomkowski et al 1992, Spere et al, 2004). The interaction of shyness and language is a fundamental part of the definition of shyness, such that an element of shyness includes talking less or being generally quieter (Buss & Plomin, 1984, Crozier, 1995, Coplan & Evans, 2009). Not surprisingly, research has supported this phenomenon, demonstrating that shy children tend to speak less in unfamiliar situations (Asendorpf & Meier,

1993). In particular, shyness is relevant for language because it could hinder a child's ability to effectively interact with others.

Certain contexts may hinder shy children's language use more than others due to their reticence in novel situations, especially those that are social in nature and directly tap traits of shyness (see also Smith Watts et al., 2014). Research supports this idea, suggesting that shyness negatively influences children's performance on language assessments or tests in a variety of settings. Hilton and Westermann (2017) show that in an experimental setting, children who are shyer are less likely to identify the correct novel object as a referent of a novel label when compared to less shy children. Additionally, shyer children tend to score lower on receptive and expressive language tasks compared to sociable children (Spere et al, 2004). Furthermore, there is evidence that shyer children typically speak less, with shorter utterances, and perform poorly on language assessments, especially in expressive language assessments (Evans, 1993). However, research is unclear if these findings stem from the socially demanding circumstances of the language assessment (e.g., the constraints of the task or environment), or from poorer language skills since no studies have examined multiple types of assessments with social interaction as the key element of examination. Additionally, while most temperament work has been done with children in preschool or older, language learning begins occurring much earlier, and examining younger populations may be beneficial.

One hypothesis, attempting to explain the relationship between shyness and language, is called the "I know it but won't say it" model (Coplan & Evans, 2009). This hypothesis states that the reason shyer children struggle to perform on language tasks is not because they are linguistically incapable, but rather because their shyness inhibits them from demonstrating their knowledge. In support of this theory is the anxiety-performance hypothesis. The anxiety-

performance hypothesis is based on evidence that shyness occurs frequently in novel environments, in evaluative settings, and among unfamiliar individuals (Asendorpf, 1989). That is, the conditions of language assessment, such as social demands, may heighten children's shyness and thus influence performance on language tests (Crozier & Hostettler, 2003). Crozier and Hostettler (2003) tested this phenomenon and found that shy children performed much better on vocabulary tests in a group environment rather than in a one-on-one, face-to-face condition. This evidence – as well as evidence that shy children perform better on receptive language tasks compared to expressive language tasks (Hilton & Westermann, 2017; Spere et al, 2004; Evans 1993, Spere & Evans, 2009) – indicates that less social tasks may lead to better performance for shy children. Thus, it is possible that shyness does not inhibit language development, but rather that social elements prevent them from demonstrating their linguistic abilities within the frame of assessments and tests.

Variability across types of language assessments

Accurate assessment of speech and communication disorders is vital to an individual's success and clinician's work. However, assessments vary greatly in the degree of social interaction required, which may differentially impact performance and the validity of such assessments. For instance, current standardized language assessments and tests can range from measurement of expressive language skills (e.g. EVT; Williams, 2019) to receptive language skills (e.g. PPVT; Dunn, 1997) to reports of vocabulary size (e.g. MCDI; Fenson et al., 1994), or even novel word learning abilities (e.g. QUILS; Golinkoff et al., 2017). In addition, language assessment in the clinic and lab often utilize speech samples and naturalistic conversations to support their assessment. Some of these formats require lots of social interaction (i.e., talking, reaching, and more) between the child and administrator which likely interact with a child's

temperament and thus, could affect that child's performance on the task. That is, since shy children feel apprehensive towards unfamiliar situations (Putnam et al, 2006), they may be less likely to socially interact with the administrator of a language assessment to respond to the prompt, even if they have strong language skills. A child's shyness (not just their vocabulary knowledge) may interact with the social demands of a task at hand to predict performance.

One method that is particularly low in social demands (and thus, possibly less influenced by shyness) are looking-based tasks. The most well-known example of this is the Intermodal Preferential Looking procedure (IPLP; Golinkoff et al., 1987). The IPLP assesses the emergence of vocabulary without requiring children to indicate their understanding through a purposeful action. In the IPLP, children are shown two images side-by-side and are told a word which matches one of the images. Then, their eye movement is recorded to determine which image they looked at for a longer period of time. This method measures language comprehension without requiring the child to speak or interact with the administrator (Golinkoff et al., 1987). Several studies demonstrate that IPLP and similar looking-based tasks reliably measures vocabulary comprehension across a variety of ages and demographic groups (Golinkoff et al, 2013; Behrend, 1990; Houston-Price et al 2007; Robinson et al, 2000). However, there are still some concerns about drop-out rates of children (Segal et al, 2021; Klein-Radukic & Zmyj, 2015). Indeed, Frank et al (2016) found that significantly more trials were dropped when children were tested in an eye-tracker compared to a storybook condition. This suggests that testing children's known comprehension, especially in looking-based tests, is far from a perfect metric. Importantly, little is known about how temperament might influence performance.

Many methods of vocabulary tests require more social interaction from the children compared to looking-based tasks. For example, a pointing or reaching method requires children

to point to or select an item from a variety of options that matches the linguistic stimulus, or a word. An example of this is the Peabody Picture Vocabulary Test (PPVT), (PPVT-R; Dunn, 1997), which requires children to point to images in a booklet. Many standardized lab-based word learning tasks also require similar pointing or selecting behaviors (Heibeck and Markman, 1987; Stelmachowicz et al, 2004). Because shyer children struggle with interactive behaviors (Putnam et al, 2006), receptive language tasks or assessments may be more difficult for them compared to less shy children (see also Hilton & Westerman, 2017). Indeed, Gurteen and colleagues (2011) tested children ages 13- to 17-month-year-olds in two different tasks, preferential looking versus preferential reaching, in word comprehension and novel word learning and found that children performed well at the preferential looking measure, but more poorly at the preferential reaching task. This suggests that the type of task matters, though how temperament plays into such differences is unknown.

A third, more socially demanding, approach for capturing children's linguistic abilities is through productive or expressive means that require children to verbalize their response. Expressive language skills can be measured through assessments such as the Test of Auditory Analysis Skills (TAAS; Rosner, 1979), Expressive One Word Picture Vocabulary Test (EOWPVT; Expressive One Word Picture Vocabulary Test, 2018), or through naturalistic observation of children's conversation and speech with adults. Comparatively, research shows that shy children tend to perform worse on expressive language tasks than receptive language tasks (Hilton & Westermann, 2017; Spere et al, 2004; Evans 1993, Spere & Evans, 2009).

Because linguistic tests are usually performed face-to-face and require social interaction with people unfamiliar to a child, the social nature of the environment could interact with children's shyness to hinder their performance. Thus, when interpreting the results of language

tasks, it is essential to differentiate between measuring children's competence and performance (Crozier, 1997). To understand this problem, Crozier (1997) recommends examining performance across a variety of situations or testing conditions that vary in social demands. Thus, the aim of the current study is to address this concern.

Current Study

The goal of this study is to examine the influence of shyness on performance on language assessment tasks that vary in degrees of social interaction, but are nearly identical in stimuli, setting, and cognitive demands. To do this, children's degree of shyness was compared with their accuracy across three different language tasks that vary in the social interaction required: a looking task, pointing task, and production language task. The hypothesis was that even after accounting for age effects, accuracy on language tasks will be driven by shyness. That is, that there will be a negative correlation between shyness and accuracy on all tasks, but the strength of the correlation will vary based on the task such that the relationship will be stronger for more socially demanding tasks (i.e., require children to verbally interact with the experimenter) and weaker for looking/reaching tasks. In addition, the prediction was that less shy children will be equally accurate across all tasks.

Methods

Participants

Participants were parents and their children, ages 17-to-42 months, who were recruited online. 122 monolingual English-speaking children participated and were primarily from the central southern United States. See Table I for demographic information. Each participant was compensated with a $20 Amazon gift card. An additional 5 children were dropped from the analysis for being bilingual (n = 3) or due to internet connectivity issues (n = 2).

Materials/Stimuli

Word/item list

A list of 24 nouns were used for the vocabulary tests, all taken from "Wordbank" (Frank et al, 2017). The words selected had the highest probability of being known by children at 17 months and had a typical or universal image which corresponded with it (e.g., the word "Momma" does not have a universal image, while "shoe" might). Approximately five images of each target word were found online. These images were normed by members of the lab (n = 14) blind to the study goals, who were instructed to select the closest image to the target word, keeping in mind that the pictures were for children, or note none of the options and explain why. Images with mixed responses were then re-assessed to ensure that they were "normal" for that target word and appropriate for children. These word-image pairs were then used in the final data sheets. Final images can be found in Appendix 1 and on OSF: https://osf.io/hwrma/?view_only=905f836347114826a54e57d6d1c10cf1 .

Temperament

The Early Child Behavior Questionnaire (ECBQ) - short form (Putnam et al, 2006) was used to measure temperament. The ECBQ is designed to capture parent-reported temperament for children from 12-36 months old. The questions ask parents to rate how often their children perform a certain behavior on a scale of 1 to 7. Responses are clustered according to three temperament dimensions: effortful control (EC), negative affect (NA), and surgency (S). Additionally, within these dimensions, 18 sub-dimensions are measured: activity level/energy (S), attentional shifting (EC), attentional focusing (EC), cuddliness (EC), discomfort (NA), fear (NA), frustration (NA), high-intensity pleasure (S), impulsivity (S), inhibitory control (EC), low-intensity pleasure (EC), motor activation (NA), perceptual sensitivity (NA), positive anticipation

(S), sadness (NA), shyness (NA), sociability (S), and soothability (NA). The sub-dimension of shyness was the focus for the current study. The ECBQ-short form includes a total of 107 items, of which 5 correspond to the shyness subscale. An example of an item on the ECBQ is: "when approached by an unfamiliar person in a public place, how often did your children pull back and avoid the person?" (Putnam et al, 2006). Research scored the questionnaire according to the standard practice (Putnam et al, 2006) and children were given a score from 1 to 7, with 1 being the least shy and 7 being the shyest. Data from the other dimensions can be found on OSF.

Vocabulary

The MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1994) was used to assess children's vocabulary. Caregivers completed either the Words and Sentences (for children 17-30-months; Fenson et al., 1994) or MCDI-III (for children 30-37-months; Dale & Fenson, 1996) depending on their child's age. Because different vocabulary metrics were used because of the wide age range and each has their own norms, instead of raw vocabulary size, each child's normative vocabulary percentile score was used in the analyses. Because vocabulary percentile is also used to assess children's language delays, such as if they are a late talker (< 20^{th} percentile), the inclusion here can also give insight on how relative vocabulary abilities for one's age impact performance on these tasks.

Procedure

Parents first completed the consent form, the ECBQ (Putnam et al, 2006), the MCDI, and demographic information via a web link within a few days of the language tasks (M = 1.45 days, SD = 3.77). The language tasks were conducted via Zoom through a PowerPoint presentation. The Zoom meeting was recorded, and children's responses were coded offline by research assistants blind to the hypothesis and child's temperament. Children were seated in their parent's

lap or on their own and parents were instructed to adjust the camera as needed to center on the child. Parents were instructed to remain silent and only offer neutral redirection if needed.

Three different types of tasks were given to the participants, each which measured vocabulary but varied in social elements. From requiring the most sociability to requiring the least sociability, the three tasks were: production task, a pointing task, and a looking task. In the production task, the experimenter directly asked the child to verbally name objects by showing a single image on the screen, and asking, “what is this?” In the pointing task, the child was instructed to point or reach for an image on the screen. Here, each trial contained two different images on either side of the screen, and the experimenter asked, “Can you point to the (target word)?” In the looking task, the child was instructed to simply look at the correct object. Each trial contained two images on either side of the screen and the experimenter prompted, “Look at the (target word).” Prompts were given immediately after the image appeared on the screen.

The order of tasks was randomized, and each participant was given each task once. Each task began with 1-2 warm-up trials in which children were offered correction and praise as needed: 1 trial for the production task and two warm-up trials for pointing and looking (one warm-up target per side of the screen). This was followed with six test trials with no praise/correction. Additionally, for each trial, experimenters only prompted the participant twice and then moved on, even if they did not respond. Orders of words and word pairs as well as task order were randomized.

Coding

Different coding was used for each task. For each task, research assistants were blind to the hypothesis and child’s temperament and achieved 90% reliability on a set of practice children prior to beginning to code. In all cases, children whose response matched the target were coded

as correct. Any response that was not to the target was marked as "not correct"; this could include a choice to the foil item, an ambiguous answer, or a lack of a response all together. The latter no responses were included in the analysis (vs. being dropped from analysis) as would be done in standard language assessments as a lack of response may also indicate shyness (i.e. a reticence to reply).

Production

The production task was coded for the correct number of phonemes that the children used per response to the prompt. To do this, coders recorded which phonemes were said correctly or incorrectly (or not at all) within their prompt. Misarticulation or accent variations of the correct phoneme was still considered correct. For example, if the correct response was "bee" and child's response was a southern-accented "beh", it was marked as correct, but if the child produced "bug" then the response would be considered 50% correct. A lack of a response was marked as 0% correct. Only 8.6% of the not-correct responses were from non-responses rather than incorrect phonemes. 68% of children were reliability coded with 92.47% agreement between coders. Any discrepancies were settled by a 3rd coder.

Pointing

To code the pointing trials, research assistants determined children's response by indicating which item they pointed to in response to the prompt. While some children pointed to a single item, other children's points were ambiguous or appeared to have changed their mind. Therefore, children's final response was taken as their choices. If the choice was to the foil or it was unclear which item was the children's final choice, it was marked as not correct. 57% of trials marked as not correct were due to an ambiguous point or lack of response. 69% of children

were reliability coded with 89.96% agreement between coders. Any discrepancies were settled by a 3rd coder.

Looking

For the looking trials, a frame-by-frame coding was used using ELAN. Starting at the end of the final word of the prompt, the coders recorded the direction of the participant's gaze for each frame (every 33ms). That is, they recorded if the participant's gaze was to the left, to the right, in between the two, or looking away from the screen. Two metrics were extracted from this: total looking time to the target on each trial and moment-to-moment gaze trajectories across each time bin (each 33ms frame segment) within a trial. The percentage of total time a child spent looking at the target on each trial was calculated for their overall accuracy. For moment-to-moment looking, average time children looked to the target in each time bin (each 33ms frame segment) were averaged across all 6 trials. 13% of children were reliability coded with 85.31% agreement between coders. Any discrepancies were settled by a 3rd coder.

Analysis

This study sought to understand the differences between children's performance on three language tasks which differed in their methodological design (e.g., level of sociability required) and their degree of shyness. The hypothesis was that accuracy across tasks would depend on children's degree of shyness; that is, shyness would predict children's performance such that shyer children performed more poorly on social tasks than non-social tasks. To test this hypothesis, two sets of analyses were run. All children who completed at least one trial in a task were included, and all trials in which the child was visibly present and/or could be heard were included in the analysis. First, bivariate correlations between shyness, age, vocabulary percentile, and task accuracy were run to give an overview of relations between variables. Second and most

importantly, mixed model regressions were used to assess trial-by-trial accuracy as predicted by a child's shyness, age, and normative vocabulary. Trial number was included as a co-variate to account for possible changes in accuracy as shy children warmed up to the task and experimenter. The first model included "task" as a predictor in order to determine if there were differences across the tasks. As described below, this was significant and thus, follow-up models examined the tasks individually. Fixed factors in both regressions included shyness score as measured on the ECBQ (centered), age (in days, centered), vocabulary percentile (normative vocabulary rank), and trial number. Possible random intercepts included participant and item; the best fitting model in most cases included both random effects. For visualization purposes only, participants were separated into a shy and non-shy group based on a median split of shyness score (median = 3.75). All analyses were preregistered and completed as planned here with additional analysis noted as exploratory. A priori power analyses estimated the sample collected here would be sufficient to detect a small-medium effect size.

Results

Overall, children were able to complete the tasks easily despite the virtual modality. In the production task, all children were presented with all six trials and across all children, were accurate 54.4% of the time (SD = .47, range = 0 - 1). In the pointing task, two children were dropped for being out of view and unable to be coded; the rest of the children were largely accurate (77.1%, SD = .32, range = 0 - 1). In the looking task, eight children were out of frame and thus, visibility of their eyes was limited. All children for whom eyes were visible were presented with all six trials. However, as further explained below, children performed rather poorly in this task with an overall accuracy of looking to the target 51.1% of the time on each trial (SD = .42, range = 0 - 1).

Correlation Analysis

First, bivariate correlations between accuracy on each task were compared to the child's age, vocabulary percentile, and shyness score. Most importantly, shyness was correlated with pointing accuracy, and production accuracy, but not target looking. See Table 2.

Regression: all tasks together

Logistic mixed model regressions were used to assess trial-by-trial accuracy across tasks. The first full model included fixed factors of shyness (centered) and task (sum coded), which interacted with age, trial number, and vocabulary percentile (all centered). The best fitting model included random intercepts of both subject and item. There were significant effects of task, age, and vocabulary percentile. There were also significant interactions. See Table 3.

Overall, shyer children tended to perform worse on the production task, while shyness did not impact performance on the looking task (see Figure 1). Importantly, age impacted each task differently (see Figure 2) such that older children performed much better on production, and slightly better on pointing, but relatively the same as younger children on looking. Finally, children with a higher normative vocabulary tended to perform much better on production, and slightly better on pointing, but relatively the same as children with lower vocabulary percentile on looking (see Figure 3). Because of these task effects, each task was further analyzed individually.

Regression: Production

A linear mixed model was used to assess the proportion of correct phonemes on each trial in the production task. Fixed factors included shyness score (centered), age (centered), vocabulary percentile (centered), and trial. Shyness was allowed to interact with the other factors. The best fitting model included a random intercept of both subject and item. There were

no significant interactions, but there were significant main effects - all children increased performance with age, with lower shyness children consistently performing better than the higher shyness children (Table 4, Figure 4). All shyness levels increased performance with vocabulary percentile, with low shyness children consistently performing better than high shyness individuals (see Figure 5). Likewise, as shyness increased, performance on production decreased (see Figure 1). The majority of "not correct" responses on these trials were incorrect phonemes. Interestingly, however, while only 8.6% of "not correct" trials were No-Responses (NR; compared to incorrect phonemes) (n = 68), 35% (n = 24 trials) of those NR trials were from less-shy children while 64.7% (n = 44) were more shy children.

Regression: Pointing

A logistic mixed model regression was used to assess trial by trial performance on the pointing task. Fixed factors included shyness score (centered), age (centered), and vocabulary percentile (centered). Shyness was allowed to interact with the other factors. The best fitting model included a random intercept of both subject and item. There was a significant effect of age, vocabulary percentile, and an interaction of shyness and age (See Table 5). The low shyness children increased accuracy with age, and children higher in shyness increased accuracy minimally across different ages (see Figure 6). Importantly, overall children were generally good at the task with most children showing above average performance.

Regression: Looking

To assess looking performance, two analyses were conducted – an overall percentage looking at the target on each trial, and a moment-by-moment analysis across each time bin. First, a linear mixed model was used to assess percent looking to the target on each trial. Fixed factors included shyness score, age, vocabulary percentile, and trial number (all centered). Shyness was

allowed to interact with the other factors. The best fitting model included a random intercept of both subject and item. There was a marginal effect of shyness, but no other significant effects (see Table 6).

Second, a logistic mixed model was used to assess bin-by-bin looking to the target. This model included fixed factors of shyness, age, vocabulary percentile, trial number, time bin (each 33ms frame), and the interactions. Subject and item were included as random intercepts. There were significant main effect of time bin (see Table 7). Over the course of the trial, children initially looked toward the target, but as the trial progressed all children looked to the target less (see Figure 7). Notably, children were overall quite poor at this task, looking to the target no more than 60% of the time, and on average at chance.

Taken together, these results demonstrate that shyness, age, and vocabulary percentile all influence accuracy on comprehension tasks. Most importantly, there were significant differences in how all three tasks interacted with the shyness level of the children. In general, children performed the best on pointing tasks, across all levels of shyness. All children performed poorly on looking tasks, despite their level of shyness. Comparatively, shyer children performed worse than less shy children on the production task, but those same shy children generally performed worse on the production task compared to the looking task.

Discussion

The main purpose of this study was to understand how shyness affects children's performance on language assessment tasks which differ in the amount of social interaction required of the child. Consistent with prior research (Spere & Evans, 2009, Hilton & Westermann, 2017, Sphere, 2004), our findings indicated that shyer children performed better on the pointing (receptive language task) rather than the production task (expressive). Importantly,

the current study tested young children (ages 17-37 months) at an early, crucial period in their language development. Our original hypothesis predicted that shyness would be negatively associated with accuracy on all tasks, but that shyer children would also be worse at more social tasks compared to non-social tasks. This was not entirely supported by our results; specifically, all children's performance (regardless of shyness) on the looking task (least social) was generally the lowest of all three. However, the results demonstrated that for the pointing task (middle social) and the production task (most social), the interaction of the social task element and shyness may have impacted children's performance. Each task is further discussed below.

Production

The findings suggest that, even after accounting for effects of age and vocabulary knowledge, shyness negatively impacts accuracy on the production task. Age and vocabulary knowledge significantly affected children's performance, such that older children/children with a higher vocabulary were much better than younger children/children with lower vocabulary (Figure 4, Figure 5). Even when accounting for these factors, it was evident that shyer children were significantly less accurate than less shy children on the production task, possibly because of the high level of social interaction requiring the child to speak. While shyer children seemed slightly more likely to give a no response compared to an incorrect word, no response trials constituted a low percentage of the "not correct" trials. Thus, the low error rate makes it difficult to draw firm conclusions.

Pointing

The results demonstrate that the children participating in the study generally performed well on the pointing task, regardless of shyness, age, or vocabulary percentile. Most significantly, high shyness children did well across development, at all ages. However, younger

children were impacted if they were lower in shyness, but shyness did not impact older children as much. That is, as children develop and get older, even the shyest children did not appear to view pointing as overly anxiety-inducing. This suggests that age may interact with temperament to predict vocabulary and that medium-social tasks may still be appropriate for slightly older children regardless of temperament.

Looking

The findings suggest that all participants struggled with the looking task, generally performing about 50% accurate regardless of shyness, age, or vocabulary percentile (Figure 1, Figure 2, Figure 3). That is, shyness did not have an impact on children's performance. This could be an indicator that the low level of interaction (i.e. no social interaction was required of the children, besides switching their gaze) did not create anxiety in shy children and thus inhibit their performance. Although all children performed poorly, shyer children did not perform *more* poorly (in fact, shyer children even performed slightly better). Perhaps this was because there was no social interaction to distract them. Instead, the low performance must have another cause.

Across the span of the looking trial, children generally started off by looking towards the target stimuli approximately 65% of the time, and then began to look away from the target quickly thereafter (See Figure 7). This provides evidence that children may have known the correct answer, but actively chose not to focus on the target for the duration of the trial, perhaps from boredom or presence of more interesting objects elsewhere. Interestingly, Frank (2016) found that for eye-tracking experiments, children ages 1- to 2-years-old generally did not complete all the trials due to asking to stop early or by other methods of refusing to continue participation. Therefore, looking-based trials may be difficult for children to remain attentive and thus responsible for children's low accuracy. Unfortunately, due to the web-based approach,

fine-grained looking behaviors beyond side of screen were not able to be coded. Additionally, there may be some stimuli and trial impacts that could be important to examine in future work.

Production and pointing

Shy children's accuracy on pointing is significantly greater than shy children's accuracy on production. While the shyest children (score 6 or lower on the ECBQ) generally performed approximately 27% accurate on production, they generally performed approximately 70% accurate on the pointing task (see Figure 1). There are several important notes about this disparity. First, shyer children consistently underperformed compared to less shy children on production at all ages (Figure 4). Second, the stimuli (i.e. pictures with corresponding words) chosen for the tasks were straightforward such that typically developing children, at least 17 months of age, had a high probability of knowing the answer (Fenson et al., 1994; Frank et al, 2017). Further, most children were well above chance on the pointing task, suggesting that they were able to understand words at that level of difficulty. Although evidence suggests that children knew the words, the shyer children were still less likely to verbally produce the correct responses.

There are other studies which demonstrate similar results: at two years old, shy children were less likely to verbal request or identify novel objects, although they had normal receptive language (Hilton & Westermann, 2017), and at four years old, shyer children scored much lower on expressive language tasks compared to less shy children (Spere at al, 2004). Most significantly, Spere and Evans (2009), found that at the kindergarten grade level, children with extreme levels of shyness were more strongly correlated with poor literacy and reading skills. The current study found a similar effect in Figure 1, suggesting that shyness impacts expressive language even at 17 months.

One possible theoretical explanation of these findings is the "I know it but won't say it" model (Coplan & Evans, 2009). In the current study, shy children performed well on the pointing task, but were unable to *produce* the correct word, this demonstrates that they were not incapable of answering due to their knowledge, but rather, they were likely inhibited (i.e. by shyness) to produce the word. Since the production task required the child to talk to the experimenter, rather than simply gesture, this socially interactive component may have triggered their anxiety and thus inhibited them from producing the correct response (Asendorpf, 1989). The results are also consistent with other theories, including the hypothesis that shy children are less likely to take risks (Coplan & Evans, 2009) and the anxiety-performance hypothesis (Asendorpf, 1989). The "bold is better" model suggests that higher sociability gives children a linguistic advantage rather than shyness inhibiting them. However, this theory does not explain why shy children performed significantly low on production tasks (Figure 1, Figure 4, Figure 5).

Limitations

This study has some potential limitations. Firstly, the study was conducted over Zoom. Virtual interactions could have differential effects on attention (McClure et al, 2017) and responsiveness from the experimenter (Strouse et al, 2018) compared to in-person testing. Therefore, these other factors could influence children's performance alongside shyness. However, there is some evidence that using teletherapy in speech-language therapy can have positive outcomes for individuals with language and communication disorders, although the evidence is mainly based on particular intervention programs (Constantinescu et al., 2014, Grogan-Johnson, 2010, Lee, Hall, & Sancibrian, 2017). According to the study by Tambyraja and colleagues (2021), 60% of responding SLPs used teletherapy during the COVID-19 pandemic, and most of them used Zoom as a virtual platform. Although more research needs to

be done about teletherapy and its efficacy, there is promising evidence suggesting it can have positive outcomes for children (Wales et al, 2017). As such, examining the efficacy of language assessment online is relevant.

Another potential limitation of this study is that the participants could have been more diverse. In our sample, 72% of the participants were white, and the average income of families was $75,000. Additionally, most of the participants were from the south-central US, and all of them were from the US. Children of lower SES tend to score lower on language assessments testing a variety of different language skills (Noble et al. 2005), as well children of minority ethnicities (Basit et al, 2015). Thus, it is important to have a diverse sample of participants to account for these potential differences.

Another note regarding the study is that the different methodologies used in the three language tasks had additional differences outside of the amount of sociability required between the experimenter and participant. Mainly, the production task measured expressive language, while the pointing and looking tasks measured receptive language. That is, there were also different cognitive demands. Therefore, this difference in methodology may have affected children's performance. Nonetheless, such differences in cognitive demands are impossible to dissect from the social elements of the task and would likely have impacted children equally despite their degree of shyness.

Conclusion

Overall, our results suggest that shyness does, in fact, impact children's performance on different language tasks. Because of this, speech pathologists, psychologists, educators, researchers, and others assessing language skills might take each individual child's shyness into consideration. These professionals should be aware that shy children may exceptionally struggle

with production assessments. In this study, shyness minimally impacted the pointing task, except for the youngest, shyest participants who performed below chance. Therefore, the results demonstrate that pointing assessments are the least biased against shy children and easier for them despite their social deficit.

Altogether, it may be important for future studies to continue to examine how to best close the gap between more shy children and less shy children's language skills. Examining how shyness impacts children's performance on standardized language assessments may be important to demonstrate specific language assessments which are most suitable for shy children. Specifically, shyness may affect children's language abilities and development, and it is important to continue to investigate how and why.

Acknowledgements

I would like to express my gratitude towards the Learning, Language, and Development Lab and Psychology Department at Oklahoma State University for providing support for this project. I would also like to acknowledge my friends and family who continuously encouraged me throughout the process.

Data Availability Statement

All materials and data for the study can be found on OSF: https://osf.io/hwrma/?view_only=905f836347114826a54e57d6d1c10cf1

552 **References**

553 Asendorpf, J. B. (1990). Development of inhibition during childhood: Evidence for situational
554 specifity and two-factor model. *Developmental Psychology*, *26,* 721–730.
555 Asendorpf, J. B., & Meier, G. H. (1993). Personality effects on children's speech in everyday
556 life: Sociability mediated exposure and shyness-mediated reactivity to social situations.
557 *Journal of Personality and Social Psychology*, *64,* 1072–1083.
558 Basit, Hughes, A., Iqbal, Z., & Cooper, J. (2015). The influence of socio-economic status and
559 ethnicity on speech and language development. *International Journal of Early Years*
560 *Education*, *23*(1), 115–133. https://doi.org/10.1080/09669760.2014.973838
561 Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label
562 familiar and novel events. *Child Development*, *61*, 681–696.
563 Buss, A. H., & Plomin, R. A. (1984). Temperament: Early developing personality traits.
564 Hillsdale, NJ: Erlbaum.
565 Briggs. (1988). Shyness: Introversion or neuroticism? *Journal of Research in Personality, 22*(3),
566 290–307. https://doi.org/10.1016/0092-6566(88)90031-1
567 Constantinescu, G., Waite, M., Dornan, D., Rushbrooke, E., Brown, J., McGovern, J., Ryan, M.,
568 & Hill, A. (2014). A pilot study of telepractice delivery for teaching listening and spoken
569 language to children with hearing loss. *Journal of telemedicine and telecare*, *20*(3), 135–
570 140. https://doi.org/10.1177/1357633X14528443
571 Coplan, R.J. & Evans, M.A. (2009). At a loss for words? Introduction to the special issue on
572 shyness and language in childhood. *Infant and Child Development*, *18*(3), 211-215.
573 https://doi.org/10.1002/icd.620
574 Crozier. (1995), Shyness and self-esteem in middle childhood. *British Journal of Educational*

Psychology, *65*, 85-95. https://doi.org/10.1111/j.2044-8279.1995.tb01133.x

Crozier. (1997). *Individual learners: Personality differences in education*. Routledge.

Crozier, & Hostettler, K. (2003), The influence of shyness on children's test performance. *British Journal of Educational Psychology*, *73*, 317-328. https://doi.org/10.1348/000709903322275858

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125–127. https://doi.org/10.3758/BF03203646

Dixon, & Shore, C. (1997). Temperamental predictors of linguistic style during multiword acquisition. *Infant Behavior & Development*, *20*(1), 99–103. https://doi.org/10.1016/S0163-6383(97)90065-5

Dunn, Lloyd M. (1997). *PPVT-III : Peabody picture vocabulary test.* Circle Pines, MN :American Guidance Service.

Eggum, Eisenberg, N., Spinrad, T. L., Reiser, M., Gaertner, B. M., Sallquist, J., & Smith, C. L. (2009). Development of shyness: Relations with children's fearfulness, sex, and maternal behavior. *Infancy*, *14*(3), 325–345. https://doi.org/10.1080/15250000902839971

Emde, Hewitt, J. K., & Kagan, J. (2001). Infancy to Early Childhood: Genetic and Environmental Influences on Developmental Change. Oxford University Press, Incorporated.

Expressive One-Word Picture Vocabulary Test. (2018). In *Encyclopedia of Clinical Neuropsychology*.

596 Evans, M. A. (1993). Communicative competence as a dimension of shyness. In *K. H. Rubin &*
597 *J. B. Asendorpf (Eds), Social Withdrawal, Inhibition and Shyness in Childhood*. Hillsdale,
598 NJ: Erlbaum
599 Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., Pethick, S., . . . Stiles, J. (1994). Variability in
600 early communicative development. *Monographs of the Society for Research in Child*
601 *Development*, *59*(5), I-185.
602 Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open
603 repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–
604 694. https://doi.org/10.1017/S0305000916000209
605 Frank, Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using Tablets to
606 Collect Data From Young Children. *Journal of Cognition and Development*, *17*(1), 1–17.
607 https://doi.org/10.1080/15248372.2015.1061528
608 Golinkoff, R.M., de Villiers, J., Hirsh-Pasek, K., Iglesias, A., & Wilson, M.S. (2017). *User's*
609 *manual for the Quick Interactive Language Screener™ (QUILS™): A measure of*
610 *vocabulary, syntax, and language acquisition skills in young children.* Baltimore: Brookes
611 Publishing.
612 Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and
613 syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23-45.
614 doi:10.1017/S030500090001271X
615 Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the
616 intermodal preferential looking paradigm to study language acquisition: What have we
617 learned? *Perspectives on Psychological Science*, *8*(3), 316–339.
618 https://doi.org/10.1177/1745691613484936

Grogan-Johnson, S., Alvares, R., Rowan, L., & Creaghead, N. (2010). A pilot study comparing the effectiveness of speech language therapy provided by telemedicine with conventional on-site therapy. *Journal of telemedicine and telecare*, *16*(3), 134–139. https://doi.org/10.1258/jtt.2009.090608

Gurteen, Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217. https://doi.org/10.1016/j.jecp.2010.12.001

Heibeck, & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*(4), 1021–1034. https://doi.org/10.1111/j.1467-8624.1987.tb01438.x

Hilton, M., & Westermann, G. (2017). The effect of shyness on children's formation and retention of novel word–object mappings. *Journal of Child Language*, *44*(6), 1394–1412. https://doi.org/10.1017/S030500091600057X

Hofmann, DiBartolo, P. M.. (2010). *Social anxiety: clinical, developmental, and social perspectives* (2nd ed.). Academic Press/Elsevier.

Houston-Price, C., Mather, E., Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behavior in a preferential looking task. *Journal of Child Language*, *34*, 701–724.

Klein-Radukic, S., & Zmyj, N. (2015). Dropout in looking time studies: The role of infants' temperament and cognitive developmental status. *Infant Behavior & Development*, *41*, 142–153. https://doi.org/10.1016/j.infbeh.2015.10.001

Lee, S., Hall, B., & Sancibrian, S. (2017). Feasibility of a Supplemental Phonological Awareness

642 Intervention via Telepractice for Children with Hearing Loss: A Preliminary
643 Study. *International journal of telerehabilitation*, *9*(1), 23–38.
644 https://doi.org/10.5195/ijt.2017.6216
645 McClure, Chentsova-Dutton, Y. E., Holochwost, S. J., Parrott, W. G., & Barr, R. (2018). Look at
646 that! Video chat and joint visual attention development among babies and toddlers. *Child*
647 *Development*, *89*(1), 27–36. https://doi.org/10.1111/cdev.12833
648 Nash, & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of*
649 *Speech, Language, and Hearing Research*, *48*(2), 439–458. https://doi.org/10.1044/1092-
650 4388(2005/030)
651 Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of
652 socioeconomic status in kindergarten children. *Developmental Science*, *8*(1), 74–87.
653 Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects
654 of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and*
655 *Development, 29*(3), 386–401. https://doi.org/10.1016/j.infbeh.2006.01.004
656 Putnam & Rothbart, M.K. (2006). Development of short and very short forms of the
657 Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102-
658 112. https://doi.org/10.1207/s15327752jpa8701_09
659 Usai, Garello, V., & Viterbori, P. (2009). Temperamental profiles and linguistic development:
660 Differences in the quality of linguistic production in relation to temperament in children of
661 28 months. *Infant Behavior & Development*, *32*(3), 322–330.
662 https://doi.org/10.1016/j.infbeh.2009.04.001
663 Robinson, C. W., Shore, W. J., Hull Smith, P., Martinelli, L. (July 2000). Developmental
664 differences in language comprehension: What 22-month-olds know when their parents are

not sure. Poster presented at the International Conference on Infant Studies, Brighton, England.

Rosner, J. (1979). *Test of Auditory Skills.* Academic Therapy Publications.

Segal, S. C., Marquis, A. R., & Moulson, M. C. (2021). Are our samples representative? Understanding whether temperament influences infant dropout rates at 3 and 7 months. *Infant Behavior & Development, 65,* 101630. https://doi.org/10.1016/j.infbeh.2021.101630

Slomkowski, C., Nelson, K., Dunn, J., & Plomin, R. (1992). Temperament and language: relations from toddlerhood to middle childhood. *Developmental Psychology*, 2*8*(6), 1090–1095. https://doi.org/10.1037/0012-1649.28.6.1090

Spere, K., Schmidt, L., Theall-Honey, L., & Martin-Chang, S. (2004). Expressive and receptive language skills of temperamentally shy preschoolers. *Infant and Child Development*, *13*(2), 123–133. https://doi.org/10.1002/icd.345

Spere, K., & Evans, M. A. (2009). Shyness as a continuous dimension and emergent literacy in young children: is there a relation? *Infant and Child Development*, *18*(4), 216–237. https://doi.org/10.1002/icd.621

Stelmachowicz, Pittman, A. L., Hoover, B. M., & Lewis, D. E. (2004). Novel-word learning in children with normal hearing and hearing loss. *Ear and Hearing*, *25*(1), 47–56. https://doi.org/10.1097/01.AUD.0000111258.98509.DE

Strouse, Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology*, *166*, 310–326. https://doi.org/10.1016/j.jecp.2017.09.005

Wales, Skinner, L., & Hayman, M. (2017). The efficacy of telehealth-delivered speech and language intervention for primary school-age children: A systematic review. *International*

Journal of Telerehabilitation, *9*(1), 55–70. https://doi.org/10.5195/ijt.2017.6219

Williams, K. T. (2019). *Expressive Vocabulary Test* (3rd ed.) [Measurement instrument]. Bloomington, MN: NCS Pearson

Tables

Table 1

Demographic characteristics of the sample.

	Mean	SD	Range
Child Age (Mo; Days)	28; 6	6; 14	17; 2 - 43; 4
Child Gender	N = 63 female		
Child Vocab			
MCDI-WS (n = 75)	197	184	679
MCDI-III (n = 46)	50.3	27.7	99
ECBQ - shyness facet	3.6	1.4	6.7
Parent Education+	6.5	3	9
Parent Income+	8.5	1	17
Race			
White	n = 88		
Black	n = 1		
Native American	n = 0		
Asian	n = 4		
Mixed race/Not listed	n = 27		
Hispanic	n = 1		

Note: Education was rank ordered with 1 as less than 7th grade and 8 as Doctoral degree; a score of 6 indicated a 4-year college degree. Income was also rank ordered with a 1 as “less than $10,000” and 9 as “more than $100,000.”

Table 2.

Correlations between each variable and task performance

	Age	Vocabulary percentile	Shyness	Target Looking	Pointing Accuracy	Production Accuracy
Age						
Vocabulary percentile	.101					
Shyness	-.182*	-.136				
Target Looking	-.038	-.105	.146			
Pointing Accuracy	.385***	.291**	-.192*	.154		
Production Accuracy	.610***	.386***	.296**	.078	.327***	

Note: Correlations between variables and task accuracy.

*p < 0.05 (2-tailed), ** p < 0.01 (2-tailed), ***p < .001 (2-tailed)

Table 3.

Accuracy across tasks by shyness

Variable	X2	df	p
Shyness	.61	1	.437
Task	106.81	2	<.001 ***
Age	54.46	1	<.001***
Vocabulary Percentile	20.56	1	<.001***
Trial	.13	1	.715
Shyness * Task	9.76	2	.008**
Shyness * Age	.12	1	.734
Shyness * Vocab Percentile	.40	1	.526
Shyness*Trial	1.18	1	.276
Task * Age	93.4	2	<.001***
Task * Vocab Percentile	40.83	2	<.001***
Task*Trial	.46	2	.796
Shyness*Task*Age	4.44	2	.109
Shyness*Task*Vocab Percentile	5.46	2	.066^{m}
Shyness*Task*Trial	2.77	2	.251

^{m}p < .10 * p < .05, **p < .01, ***p < .001

Table 4.

Regression of Trial-by-Trial Accuracy on Production

Variable	β	**t**	p
Shyness	-.05	-2.16	.033*
Age	.19	8.35	<.001***
Vocabulary Percentile	.11	5.17	<.001***
Trial	-.00	-.08	.941
Shyness*Age	.01	.22	.824
Shyness*Vocab Percentile	-.00	-.17	.869
Shyness*Trial	-.02	-1.48	.139

*p < .05, ***p < .001

Table 5.

Regression of trial-by-trial accuracy on Pointing

Variable	***β***	**z**	***p***
Shyness	-.16	-.90	.365
Age	.846	4.91	< .001***
Vocabulary Percentile	.576	3.21	.001**
Trial	.05	45	.650
Shyness * Age	-.37	-2.04	.042*
Shyness*Vocabulary percentile	.18	.98	.325
Shyness*Trial	-.13	-1.26	207

*p < .05, **p < .01, ***p < .001

Table 6.

Regression of Trial-by-Trial Accuracy on Looking

Variable	***β***	**t**	**p**
Shyness	.03	1.66	.10^{m}
Age	-.00	-.12	.908
Vocabulary Percentile	.01	-.69	.492
Trial	.00	.11	.919
Shyness*Age	.01	-.37	.714
Shyness*Vocab	.01	.56	.577
Shyness*Trial	.01	.92	.357

$^{m}p < .10$

Table 7.

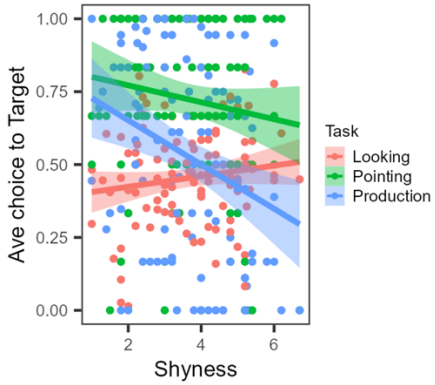
Accuracy across time bins in Looking

Variable	*β*	z	p
Shyness	.13	.73	.464
Age	-.09	-.52	.606
Vocab Percentile	-.07	-.41	.680
Trial	.06	.77	.440
Time Bin	-.18	-18.26	<.001***
Shyness*Age	.23	1.34	.181
Shyness*Vocab	-.10	-.61	.540
Shyness*Trial	.01	1.50	.134
Shyness*Bin	-.02	-1.51	.131

***p < .001

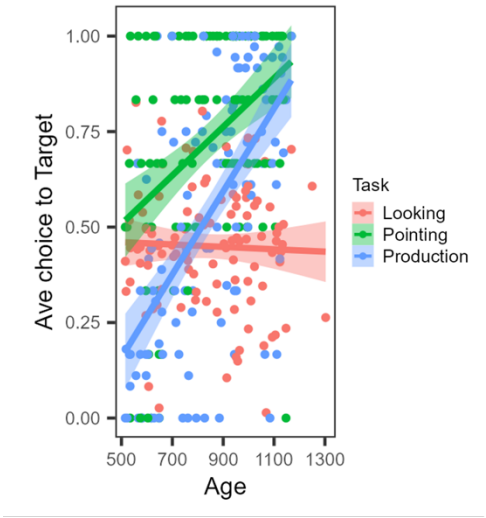
Figure legends

Figure 1: *Relation between accuracy and shyness, according to task*



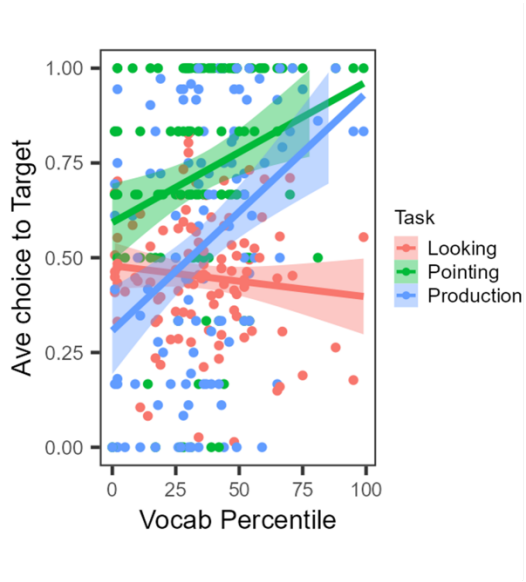
Note: Linear regression lines shown for visualization purposes only

Figure 2: *Accuracy across tasks by age*



Note: Linear regression lines shown for visualization purposes only.

Figure 3: *Accuracy on tasks by vocabulary percentile*



Note: Linear regression lines shown for visualization purposes only

Figure 4: *Accuracy on production by age*

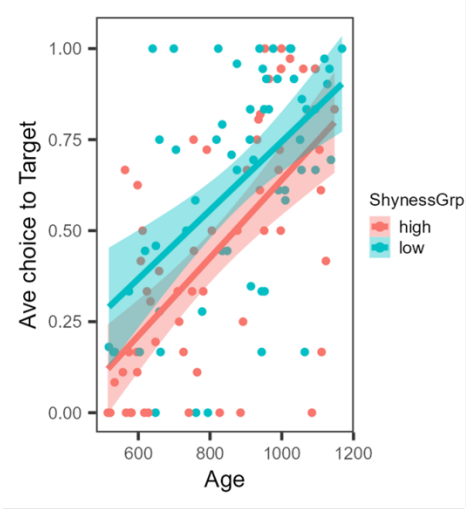


Figure 5: *Accuracy on production by vocabulary percentile*

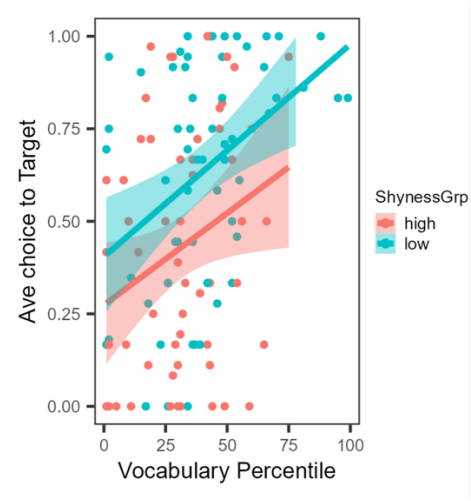


Figure 6: *Accuracy on pointing by age*

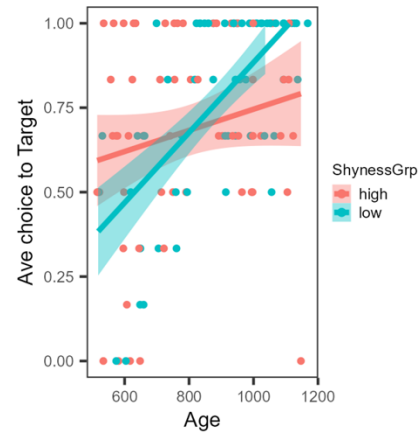


Figure 7: *Moment-by-moment looking patterns across time bin*