

Research Note

Taking Language Samples Home: Feasibility, Reliability, and Validity of Child Language Samples Conducted Remotely With Video Chat Versus In-Person

Brittany L. Manning,^a Alexandra Harpole,^a Emily M. Harriott,^a Kamila Postolowicz,^a and Elizabeth S. Norton^{a,b} 

Purpose: There has been increased interest in using telepractice for involving more diverse children in research and clinical services, as well as when in-person assessment is challenging, such as during COVID-19. Little is known, however, about the feasibility, reliability, and validity of language samples when conducted via telepractice.

Method: Child language samples from parent-child play were recorded either in person in the laboratory or via video chat at home, using parents' preferred commercially available software on their own device. Samples were transcribed and analyzed using Systematic Analysis of Language Transcripts software. Analyses compared measures between-subjects for 46 dyads who completed video chat language samples versus 16 who completed in-person samples; within-subjects analyses were conducted for a subset of 13 dyads who completed both types. Groups did not differ significantly on child age, sex, or socioeconomic status.

Results: The number of usable samples and percent of utterances with intelligible audio signal did not differ significantly for in-person versus video chat language samples. Child speech and language characteristics (including mean length of utterance, type-token ratio, number of different words, grammatical errors/omissions, and child speech intelligibility) did not differ significantly between in-person and video chat methods. This was the case for between-group analyses and within-child comparisons. Furthermore, transcription reliability (conducted on a subset of samples) was high and did not differ between in-person and video chat methods.

Conclusions: This study demonstrates that child language samples collected via video chat are largely comparable to in-person samples in terms of key speech and language measures. Best practices for maximizing data quality for using video chat language samples are provided.

Language sampling is a common assessment tool speech-language pathologists (SLPs) and researchers use to evaluate children's spontaneous, expressive language (Costanza-Smith, 2010; Hux et al., 1993). Language samples are often elicited from a child during play or story retell, providing the advantage of a more naturalistic and representative picture of the child's language use than most standardized assessments generate (Evans & Craig, 1992). This spontaneous language can be transcribed

and analyzed for various linguistic elements including syntax, morphology, fluency, semantics, and narrative skills. Parent-child interaction can also be coded from such a sample and analyzed for a variety of behaviors relevant to language, such as parent responsiveness (e.g., Tamis-LeMonda et al., 2001) or strategy use (Bullard et al., 2017). Language sampling can also serve as a more neutral or less culturally biased form of assessment for linguistically diverse children (Newkirk-Turner et al., 2016; Oetting & McDonald, 2002; Stockman, 1996) as it does not rely on stimuli that may be culturally inappropriate or unfamiliar.

Language samples are typically conducted with an examiner or a parent at a clinical site or a research laboratory; however, as the goal of language sampling is to elicit naturalistic language, the setting itself may be a disadvantage, as it may be unfamiliar to the child. Studies have indicated that children's language when at home differs in form, content, and use as compared to other settings (Kramer

^aRoxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL

^bDepartment of Medical Social Sciences and Institute for Innovations in Developmental Sciences, Northwestern University, Chicago, IL

Correspondence to Elizabeth S. Norton: enorton@northwestern.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Stephanie Borrie

Received April 23, 2020

Revision received July 2, 2020

Accepted July 26, 2020

https://doi.org/10.1044/2020_JSLHR-20-00202

Disclosure: The authors have declared that no competing interests existed at the time of publication.

et al., 1979; Larson et al., 2020). Language sampling, like most language assessments, typically requires in-person interaction, thus limiting the individuals who may benefit from clinical assessments or be included in research (Ciccia et al., 2011). Evidence-based guidelines exist for language sample elicitation (Costanza-Smith, 2010; Guo & Eisenberg, 2015) and there are well-established procedures for analyzing and reporting data from language samples (Finestack et al., 2014); however, very little research has examined whether language samples can be collected outside of traditional in-person interactions. This study aims to investigate the feasibility, validity, and reliability of a procedure for collecting language samples remotely from children and parents in their homes, using commercially available video chat software.

Limitations of Current Language Sampling Procedures and Opportunities for Telepractice

Although limitations currently exist for collecting language samples via remote methods, there is evidence that this practice would be beneficial to both clinical practice and research. In fact, the American Speech-Language-Hearing Association (ASHA) supports telepractice as a way to increase accessibility for rural and underserved communities (ASHA, 2011; Robertson, 2019). Lack of transportation or child care, economic hardship, and/or cultural beliefs may limit the accessibility of speech-language services for children in rural and urban areas (Ciccia et al., 2011; Hall et al., 1991; Verdon et al., 2011), but a telepractice approach can minimize some of these barriers (Hill & Theodoros, 2002; Hodge et al., 2019). Given that the majority of families in the United States have smartphones or devices with Internet access at home (Ryan, 2017), providing services via video chat is highly feasible for many families.

Researchers also face the challenge of including diverse participant samples. This is important for many reasons, including ensuring the generalizability of research findings (Hammer, 2011). Families of rural, minority, or low socioeconomic status backgrounds face limitations to engaging in research similar to those of receiving services, and further, may lack trust in researchers (George et al. 2014). Given that language samples are one of the most common tools used in research reported in ASHA journals (Finestack et al., 2014), the capacity to collect reliable language samples via video chat could make participating in research easier for many families and improve the quality and generalizability of research.

Research on Remote and Telepractice Speech-Language Assessment

Telepractice is already in broad use. In 2016, 60.7% of SLP members of ASHA's Telepractice/Telehealth Special Interest Group reported conducting speech-language assessment via telepractice (ASHA, 2016); however, few studies have analyzed the reliability and validity of this type of

language assessment (Taylor et al., 2014). To our knowledge, only one study has compared naturalistic language assessment between in-person and telepractice contexts. This study examined a story retell paradigm in 40 adults following brain injury (Brennan et al., 2004). Participants completed two conditions: an in-person condition where they were in the same room with the clinician and a telepractice condition where the clinician was in a different room within the same hospital. The authors found no significant differences in their outcome measure, percent of correctly identified story information, between in-person and telepractice assessments, and there was a high correlation ($r = .93$) between performance in the two conditions.

Very few studies have looked at the feasibility, validity, or reliability of telepractice speech and language assessment in children, and none to our knowledge have specifically examined the validity of language sampling. One study examined remote assessment and parent-implemented intervention in three families who had a child with fragile X syndrome. Researchers gave each family a laptop to use and collected language samples and other measures via Skype (Bullard et al., 2017). Because all their measures were remote, they did not directly examine how the data compared with in-person assessment. They did report that no families found the technology limiting or a hindrance to participating in the intervention, and suggested that further research was needed to determine the reliability and validity of language samples as a tool for assessing treatment gains.

Several studies of speech sound/articulation abilities and oromotor function in children ages 4–9 years have been carried out and suggest that agreement between scoring in-person and telepractice recordings of speech ranges from fair to good (Eriks-Brophy et al., 2008; Waite et al., 2006, 2012). For detailed assessment at the phoneme level, very high-quality recordings may be needed, yet remote data collection still holds promise for speech-language assessment.

Studies assessing child language via telepractice have mostly focused on adapting standardized screening and assessment measures to remote administration. In one study, the feasibility of telepractice speech, language, and hearing screenings was examined in children up to age 6 years via video chat (Ciccia et al., 2011). Pass/fail rates were not found to differ between in-person and video chat screenings in this small sample ($n = 10$). In another method comparison case study, seven children completed multiple standardized language assessments; one SLP administered and scored the assessments via telepractice, while another SLP was in the room with the child and scored the assessments as they were administered (Eriks-Brophy et al., 2008). The interrater reliability was 98.4%–100% across assessments, indicating that the telepractice scores were highly similar to scores obtained in person. A similar study with 25 children examined the validity of administering and scoring the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel et al., 2003) via telepractice (Waite et al., 2010). Participants were randomly assigned

to a telepractice or in-person condition, with both SLPs watching and scoring each session in real time. Raw scores did not differ significantly between the conditions and there was “very good” rater agreement ($k > .90$) for all scores. The largest reported difference was between raw scores for the concepts and following directions subtest in the telepractice ($M = 21.28$) and in-person condition ($M = 21.52$); however, this was not significant ($p = .08$). Finally, another study examined CELF-4 administration via telepractice in 23 school-age children (Sutherland et al. 2017). They found a strong correlation ($r = .96$ – 1.0) between the telepractice and in-person clinician scores, and audio or video quality was rated as poor in only one session. In all, these studies suggest that remote language assessment is promising.

Current Study

Despite the potential of remote language assessments to ease burden on participants, to our knowledge, no study has examined the feasibility, validity, and reliability of using video chat with families’ own devices and free commercial software to conduct language sampling for children. We hypothesized that overall, in-person and video chat methods would yield similar metrics, despite differences in the child’s environment, given the similarity of the parent–child play context. We collected language samples with toddlers and their parents, both in person in the laboratory and at home via video chat and examined the following research questions:

1. Feasibility: Is data quality, as indicated by the number of usable samples and percent of utterances with intelligible audio signal, similar for language samples collected via video chat as compared with traditional in-person methods?
2. Validity: Do widely used speech and language metrics including mean length of utterance (MLU), number of different words (NDW), type–token ratio (TTR), number of language errors and omissions, and percent of utterances with child speech intelligible differ between video chat and in-person samples?
3. Reliability: Does transcription reliability differ between language samples obtained via video chat versus in-person?

Method

Study Overview

The study was a mixed quasi-experimental and longitudinal design. Procedures were approved by Northwestern University’s institutional review board (IRB). Parents provided informed consent via video chat or in person. Families received monetary compensation for their time.

Study Structure and Design

This study was part of a larger research study designed to assess the effect of an app-based intervention for parents on their child’s language and validate the use

of language samples via remote and in-person methods. Each toddler was enrolled in one of two study visit structures that dictated whether they completed in-person or video chat language samples at each timepoint. Families in the mixed visit structure visited the laboratory in person twice (pre-intervention at Timepoint 1, then 6 weeks later at Timepoint 3 postintervention) to complete a language sample and standardized language assessments; they also completed language samples via video chat at midpoint in Week 3 (Timepoint 2) and at follow-up in Week 10 (Timepoint 4). The video chat structure group completed language samples via video chat at Timepoints 1, 2, and 3 (i.e., pre, Week 3 midpoint, Week 6 posttest; no follow-up). Families were given the option of enrolling in the mixed or video chat visit structure until the planned sample size of mixed visit families was enrolled; afterwards, all families were enrolled in the video chat structure. This study design allowed comparisons of the two groups pre-intervention (as all mixed structure families had in-person samples at Week 1), as well as within-subjects comparisons of the mixed structure families who had two in-person and two video chat language samples.

Participants

Families with toddlers were recruited via social media and word of mouth. Parents completed an eligibility survey with the following criteria: toddler aged 18 to 34 months, gestation at least 37 weeks, no signs of hearing loss, no other medical diagnoses that would impact language development, learning American English as a first language, and lives in the United States. Additional inclusion criteria included access to a smartphone and Internet.

The planned sample size for this pilot study was 20 children in the mixed visit structure and approximately 60 children in the video chat structure. Data were initially collected from 77 participants. One child was excluded whose video was lost due to a technical error (at Timepoint 1, central to current analyses). We excluded 14 children whose samples were transcribed but did not reach the minimum level of 50 complete (not interrupted or abandoned) and intelligible utterances at all timepoints, as this is considered a minimum for clinical and research sample validity (Eisenberg, 2001; Heilmann et al., 2010). An SLP then reviewed these transcripts and confirmed that children were primarily in proto-word or babbling stages and used few real words. Furthermore, three children from the mixed visit structure group are included in between-group but not within-group analyses, as one child’s Timepoint 4 video quality was unusable and two families were pilot participants in the mixed-visit structure who completed samples at Timepoints 1, 2, and 3 only.

Participants were diverse in terms of racial/ethnic background. Parents reported their child’s race. The mixed visit structure group included 6.3% American Indian or Alaska Native; 25.0% Black/African American, 6.3% Native Hawaiian/Pacific Islander, and 62.5% Caucasian; 18.8% were Hispanic. The video chat group included 8.7% Asian,

8.7% Black/African American, 63.0% Caucasian, and 19.6% more than one race; 6.5% were Hispanic.

Procedure

Parent Intervention

The pilot parent intervention is not a focus of the current study but is described here given the multiple measures in the mixed visit group during intervention. Families were told that they would be assigned to one of two app versions, each of which sent reminders to the parent's smartphone 5 times per day. The study app provided parents with tips for rich quality and quantity of language interactions, customized to the child's age; the control app provided only generic reminders about quantity (e.g., "Remember to talk to Jaden!"). All families in the mixed visit structure were assigned to the study app; families in the video chat structure were pseudorandomized into the study or control app, matching for demographics. Families were not told whether their app was the study app or control app, but at the end of intervention, all families were invited to use or to continue using the study app if they wished.

Language Sample Procedures

Unstructured play with the parent was used as the sampling context for both in-person and video chat language samples; parents were told to play with their child the way they normally would so that researchers could observe the child's language abilities. Examiners were instructed to collect at least 15 min of parent-child interaction and to collect 20 min or more if the child was not using much language.

In-person language samples were conducted in a sound-attenuated room in the laboratory equipped with two Panasonic PTZ HD video cameras and a Shure omnidirectional table microphone, and recorded to a single file using an Extron switcher. The toddler and parent sat at a table with a farm play set, pretend ice cream set, and/or blocks. The examiner sat in the corner of the room.

For video chat language samples, families used their own device such as a smartphone, tablet, or computer. They chose whether to use Skype, Google Hangouts, or FaceTime, which were popular software platforms that were IRB-approved for this study. (As this was a research study, HIPAA compliance was not required.) Families used their home Internet (ethernet or Wi-Fi) or cellular connection; the type and strength of the network connection varied and was not limited or set as an exclusionary criterion in order to mirror the real-world conditions of different families. The researcher recorded the interaction via screen capture in QuickTime Player software on a MacBook laptop that was connected to Wi-Fi with bandwidth of approximately 800–1,000 Mbps. In advance of the video chat session, suggestions for toys to use during the session were given to families and electronics were discouraged; however, families chose what toys worked best for them.

Language Sample Transcription, Analysis, and Reliability

All language samples were transcribed using Systematic Analysis of Language Transcripts (SALT) software (Miller & Iglesias, 2015) by trained undergraduate research assistants studying communication sciences and disorders and/or neuroscience and graduate research assistants studying clinical speech-language pathology. All research assistants were supervised by a certified SLP. All transcribers reached >80% accuracy on three consecutive practice samples before beginning transcriptions. Guidelines from the SALT manual were followed for C-unit segmentation and transcription conventions. A custom code was used for utterances that were unintelligible due to poor audio including background noise, noise from toys, noise from others in the room, and so forth. Utterances with nonlinguistic vocalizations (e.g., cries, grunts), proto-words, and/or babbling were assigned a separate code and excluded from the analyses.

To assess transcription reliability, approximately 25% of samples were double transcribed by a graduate research assistant who did not collect the original sample. Reliability was determined by dividing the number of matching words, morphemes, and codes between the two transcripts by the total words/morphemes/codes. Reliability assessed codes for unintelligibility, errors, omissions, custom codes, and fluency codes, which were not used for the remainder of the analyses. Mean reliability was 88.59%, with a range of 82%–98%. These reliability percentages are in line with previously published studies of toddler language samples (e.g., Hadley, 1998). If discrepancies were identified, the two transcribers met to reach consensus for the final transcript.

We first calculated the number of usable video chats, those that did not have significantly poor audio that precluded them from transcription, across all timepoints and participants (166 samples total). Then, the following variables were calculated using SALT for each transcript. Data quality was indicated by percent of utterances with audio signal intelligible, the percent of all utterances with no words or segments of unintelligibility due to poor audio quality, which was calculated using a custom code in SALT. Child speech and language metrics calculated with SALT were MLU in morphemes, NDW (number of different words used), TTR (number of different words/total words), number of errors (utterance, word, and overgeneralization errors), and number of omissions (omitted words and bound morphemes). These metrics were calculated by the standard SALT procedure using a transcription cut of 50 complete and intelligible utterances; thus, all samples had exactly 50 utterances analyzed. The metrics were chosen for analysis because they are commonly used in clinical practice (e.g., Pezold et al., 2020). Finally, we calculated percent of utterances with child speech intelligible as the percent of verbal utterances with usable audio that contained no unintelligible words or segments due to child speech. Consistent with our broader research goals, nonlinguistic

vocalizations, proto-words, and babbling were also excluded from this calculation.

Statistical Analyses

Nonparametric statistics were used for all analyses because dependent variables did not meet assumptions of normality. Mann–Whitney *U* tests were conducted to examine group differences between the in-person and video chat language sample groups at Timepoint 1. Wilcoxon signed-ranks tests were used to compare within-subjects differences between language samples conducted in person versus via video chat in the mixed visit structure group. A mean score was calculated for both in-person language samples (conducted at Timepoints 1 and 3) as well as for video chat language samples (conducted at Timepoints 2 and 4). For all analyses, effect sizes are reported in units of *r* (absolute values), $r = \frac{z}{\sqrt{n}}$. This can be interpreted in the same units as Pearson's *r* correlation coefficient, with effect sizes of 0.1–0.3 considered small (Fritz et al., 2012). Spearman correlations were used to assess continuous relations among measures.

Group differences were also examined for sex, age, and socioeconomic status, indicated by income-to-needs ratio. Income-to-needs-ratio was calculated as reported family annual income divided by the federal poverty criterion for a household of that size (U.S. Census Bureau, 2018). Thus, a family whose income was exactly at the federal poverty line for their household size would have a value of 1.0.

Results

Descriptive information, statistics, and effect sizes for comparisons of the in-person versus the video chat visit samples at Timepoint 1 are given in Table 1. The groups of participants in the two visit structures did not significantly differ in terms of child sex ($\chi^2 = 1.07$, $p = .30$), child age ($p = .42$), or family income-to-needs ratio ($p = .23$).

To assess feasibility and data quality, we first calculated the number of usable samples collected via video chat. Of the 166 total video chat samples collected, only

two video chat samples were unable to be transcribed due to audio; thus, 98.8% of samples collected had sufficient audio quality. The reasons for the two samples' poor quality were that one family conducted their language sample on their porch with significant background noise, and the other family moved to an area with extremely poor Internet connection. We next examined the percent of utterances with audio signal intelligible among usable samples for the two methods. Both groups had, on average, above 98% of utterances with usable audio and there was no significant difference between groups. The lower limit of the range was higher in the in-person group in both cases, although the effect size of the mean difference was small ($r = .01$). To determine whether technology differences associated with family socioeconomic status may influence these results, we examined whether family income-to-needs ratio related to percent audio signal intelligible; there was no significant correlation between these two variables ($r_s = -.08$, $p = .52$).

Next, we examined the validity of child speech and language metrics MLU, NDW, TTR language errors and omissions, and utterances with intelligible child speech between the initial in-person and video chat groups. Results revealed no significant difference between the groups on any measures, with the smallest effect sizes for NDW ($p = .90$; $r = .02$), MLU ($p = .85$; $r = .02$), number of errors ($p = .87$; $r = .02$), and percent intelligible child speech ($p = .72$; $r = .05$).

We further compared data quality and validity of speech and language measures within participants in the mixed visit structure group who completed in-person and video chat samples. Descriptive statistics and comparisons between variables of interest for these measures are given in Table 2. As in the between-group comparison, there were no significant differences between conditions for any measures and all effect sizes were small. Here, the largest effect size was for our data quality measure, percent utterances with audio signal intelligible, which was above 96% in both methods.

Next, because transcript analysis is typically the goal of language sampling, we compared transcript reliability in a subset of ~25% of samples selected randomly from

Table 1. Descriptive statistics and statistical comparisons for in-person versus video chat samples at Timepoint 1.

Variable	In-person (<i>n</i> = 16)		Video chat (<i>n</i> = 46)		<i>z</i>	<i>r</i>	<i>p</i>
	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range			
Age (months)	24.00 (4.02)	19–32	24.87 (4.24)	18–34	–0.81	.10	.42
Income:Needs ratio	4.65 (4.10)	0.61–14.69	5.66 (4.31)	0.69–19.59	–1.20	.15	.23
% utterances audio signal intelligible	99.07% (1.36%)	95.00%–100%	98.44% (3.29%)	81.75%–100%	–0.07	.01	.95
% utterances child speech intelligible	86.31% (9.03%)	68.00%–100%	84.91% (10.08%)	60.00%–100%	–0.36	.05	.72
Type–token ratio	0.48 (0.09)	0.34–0.62	0.46 (0.09)	0.25–0.69	–1.05	.13	.30
Mean length of utterance	1.69 (0.58)	1.02–2.60	1.75 (0.68)	1.02–3.58	–0.19	.02	.85
Number of different words	37.13 (11.90)	13–68	37.59 (12.63)	20–64	–0.13	.02	.90
Errors	1.06 (1.34)	0–4	1.11 (1.64)	0–7	–0.17	.02	.87
Omissions	3.63 (4.66)	0–18	3.83 (3.95)	0–19	–0.62	.08	.54

Note. All *z* values are from Mann–Whitney *U* tests. Effect size measured in units of *r*.

Table 2. Descriptive statistics and comparisons within-subjects for the mixed visit structure group ($n = 13$): In-person (Timepoints 1 + 3) versus video chat language samples (Timepoints 2 + 4).

Variable	In-person		Video chat		z	r	p
	M (SD)	Range	M (SD)	Range			
% utterances audio signal intelligible	98.78% (1.15%)	96.14%–100%	96.63% (5.15%)	84.02%–100%	-1.08	.21	.28
% utterances child speech intelligible	86.50% (7.65%)	71.00%–96.50%	84.00% (6.54%)	75.00%–99.00%	-0.98	.19	.33
Type–token ratio	0.48 (0.07)	0.38–0.59	0.47 (0.11)	0.22–0.66	-0.25	.05	.81
Mean length of utterance	1.90 (0.62)	1.13–2.98	1.90 (0.63)	1.03–3.04	-0.46	.09	.65
Number of different words	40.23 (10.42)	24.00–59.50	41.65 (13.44)	11.50–63.50	-0.67	.13	.51
Errors	1.23 (1.25)	0–4.5	1.04 (1.33)	0–4.0	-0.94	.19	.35
Omissions	4.73 (4.35)	0–15.0	5.42 (5.28)	0–15.50	-0.99	.19	.33

Note. All z values are from Mann–Whitney U tests. Effect size measured in units of r.

across all timepoints (nine in-person and 29 video chat). Mean reliability values were extremely similar across contexts (in-person $M = 88.89\%$, $SD = 5.3\%$; video chat $M = 88.50\%$, $SD = 4.3\%$) and there was no statistically significant difference between conditions ($r = .05$; $p = .76$). Finally, we confirmed that there were strong and significant ($p < .001$) expected correlations between child age and Timepoint 1 measures of MLU ($r_s = .75$), NDW ($r_s = .63$), errors ($r_s = .54$), and omissions ($r_s = .70$).

Discussion

This study is the first to report on the feasibility, validity, or reliability of remote language sampling for child language assessment. The importance of collecting high-quality assessments from families who cannot easily be assessed in person or do not feel comfortable in traditional research environments has created a need for empirical studies of remote child language assessment methods. Language samples are particularly suited for assessing children of diverse cultural and linguistic backgrounds (Oetting & McDonald, 2002; Stockman, 1996). The need for validated telepractice measures has also been magnified by the COVID-19 pandemic. Consistent with the limited previous research on assessment via telepractice, pilot data here from a diverse sample of toddlers indicated that there were no significant or meaningful differences in child speech and language measures or transcription reliability between video chat and in-person language samples. This was true for both analyses between two groups and within one group who completed language samples in person in the laboratory and via video chat.

In terms of data quality, just two of 166 samples attempted, with our 62 participants using their own devices and various types of Internet connections for video chat, had poor audio quality that precluded transcription and analysis. This rate of 1.2% unusable samples is similar to or better than previous telepractice research (1/23 or 4.3% unusable sample rate in Sutherland et al., 2017). Of usable samples, the percent of utterances with intelligible audio signal was slightly higher during in-person samples; however, the mean of the video chat samples was above 96% in

both between- and within-subject analyses. This suggests that, although researchers should be mindful of collecting a long enough sample to yield at least 50 intelligible and complete utterances, there is not substantial data loss from the video chat method in the vast majority of cases.

Child language measures were also highly comparable between methods. Group mean MLU was identical for the within-subjects comparison of two in-laboratory in-person samples with two video chat samples each collected 3–4 weeks later. For the between-subjects analysis, the video chat group mean (1.75) was slightly higher than the mean of the in-person group (1.69); small numerical differences in MLU of 0.02–0.16 exist even when the exact same language samples are transcribed/analyzed using different software (Pezold et al., 2020). The transcription reliability observed for both video chat and in-person language samples in this study also aligns with existing studies that focus on transcriptions of toddler language (Hadley, 1998).

Some limitations of this study should be considered. First, data reported here include only a parent–child play context with toddlers aged 18–34 months, which is on the younger end of the ages language samples are typically used. Older children may interact differently in a video chat or this may vary by sampling context. Second, our video chats were collected using families' own devices (various types of phones, tablets, and computers) and multiple commercial platforms, so these differ from some telepractice approaches that used high-quality clinic video systems (Brennan et al., 2004) or that provided families with equipment (e.g., Bullard et al., 2017). Third, the within-subject analyses compared data that were collected during the course of a low-intensity parent intervention; the lack of difference between the in-person samples at Timepoints 1 and 3 (pre- and postintervention) versus video chat samples at Timepoints 2 and 4 (midintervention and follow-up) suggests that this did not strongly influence our results. We encourage future studies to examine within-participant differences between video chat and in-person language samples outside of the context of an intervention study to examine this question in more detail, and to use larger samples with greater diversity to understand potential cultural differences. Our design also did not allow us to

examine differences due to the home environment and the video chat technology itself. In fact, previous studies have shown that the home environment may actually elicit more rich child language output (Kramer et al., 1979; Larson et al., 2020). Future studies should compare in-person and video chat recordings of parent–child interaction both conducted in the home to differentiate the effects of the recording method and testing environment. Finally, our percent child speech intelligibility variable was calculated somewhat differently than the traditional SALT variable. We excluded utterances with unusable audio in order to separately examine technological difficulties due to video chat. We also separately coded utterances with babbling in order to assess advancement in babbling as part of our larger study aims. Researchers and clinicians should be mindful that further validation of this measure is needed. Importantly, previous studies have shown that speech measures may be more difficult to assess using remote methods than language measures (Eriks-Brophy et al., 2008; Waite et al., 2006). A trend toward lower child speech intelligibility as calculated here was observed in the within-subjects analysis. This may have been influenced by toy selection; in the laboratory, children often played with a farm toy that elicited simple nouns that were highly intelligible, but toys at home like play-dough allowed fewer of these opportunities. Future research should examine in-home samples with standard toy sets to compare more closely with previous research. Furthermore, analysis of children’s phonological or articulation errors was not the focus of this study, but this is an important area for future research.

Overall, our pilot data suggest feasibility of video chat methods for evaluating children’s spontaneous language production. Given the promise of telepractice for improving access to speech and language therapy and research participation for underserved families and the necessity of remote assessment during the COVID-19 pandemic, researchers should continue to develop and validate remote methods for assessing child speech and language.

Best Practices and Lessons Learned for Video Chat Language Samples

Some of our team’s “lessons learned” and recommendations for video chat language samples are provided below. A full description of our video chat procedures can be found at <http://learnlab.northwestern.edu/resources>.

Planning for the Session

- It is best to talk with the parent in advance of the session to give them relevant information and set their expectations. Answering all of the parent’s questions before the session is ideal, so as not to delay the beginning of the session or have the toddler get impatient.
- Let parents know that the session should take place indoors in a quiet room with no other adults, children, or pets present.

- Discuss the types of toys that work well; toys that make a lot of noise and electronics should be avoided if possible.
- Arrange to send a “test message” via Skype or other platforms that require a username in advance of the session, to ensure that the correct person is contacted and to try to avoid technical problems at the beginning of the session.

Addressing Parent Concerns

- If a parent has concerns about their child having screen time, reassure them that the child will not be directly interacting with the screen.
- If they are concerned that their child will not stay in one place, encourage them to use a device that can be moved around with the child, like a tablet or mobile phone.

Technology

- Ensure that your choice of video chat software is compliant with relevant IRB and/or HIPAA regulations (this may vary based on institution).
- Researchers can share directions for creating an account or downloading software or plugins for video chat if the family does not already use one of the IRB-approved software programs.
- Encourage families to use ethernet or Wi-Fi rather than a cellular signal when possible. If the family has concerns about quality of the cellular or Wi-Fi signal, a test call can take place in advance. Let families know that they may incur data charges on cellular signals.
- If the parent is using a phone or tablet, placing it in a horizontal rather than vertical orientation allows for a wider field of view, so it is easier to see the child if he/she moves. Sound quality is also often better when the microphone is not up against a table/flat surface.

Acknowledgments

This work was funded by a grant from the Delaney Fund for Research in Communication (School of Communication, Northwestern University) to PI Norton and by support from the Undergraduate Research Grant Program, which is administered by Northwestern University’s Office of Undergraduate Research. The conclusions, opinions, and other statements in this publication are the authors’ and not necessarily those of the sponsoring institutions.

We acknowledge these contributors to this research: Katie Gottfred, John Lybolt, Nina Smith, Josh Holton, Ryan Gunn, Emma Baime, Camille Nuttall, Biya Ahmed, Maggie Boland, Shradha Mehta, Jade Mitchell, Skylar Ozoh, Cadence Reed-Bippen, Kiera Cook, Sean McWeeny, and Silvia Clement-Lam. We thank Pamela Hadley for her suggestions on the paper. We thank the participating families for their time.

References

- American Speech-Language-Hearing Association.** (2011). *Scope of practice*. www.asha.org/policy/SP2016-00343/
- American Speech-Language-Hearing Association.** (2016). *2016 SIG 18 telepractice survey results*. www.asha.org/uploadedFiles/ASHA/Practice_Portal/Professional_Issues/Telepractice/2016-Telepractice-Survey.pdf
- Brennan, D. M., Georgeadis, A. C., Baron, C. R., & Barker, L. M.** (2004). The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemedicine Journal and e-Health, 10*(2), 147–154. <https://doi.org/10.1089/tmj.2004.10.147>
- Bullard, L., McDuffie, A., & Abbeduto, L.** (2017). Distance delivery of a parent-implemented language intervention for young boys with fragile X syndrome. *Autism & Developmental Language Impairments, 2*. <https://doi.org/10.1177/2396941517728690>
- Ciccia, A. H., Whitford, B., Krumm, M., & McNeal, K.** (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *Journal of Telemedicine and Telecare, 17*(5), 240–244. <https://doi.org/10.1258/jtt.2011.100810>
- Costanza-Smith, A.** (2010). The clinical utility of language samples. *Perspectives on Language Learning and Education, 17*(1), 9–15. <https://doi.org/10.1044/ll17.1.9>
- Eisenberg, S.** (2001). Use of MLU for identifying language impairment in preschool children: A Review. *American Journal of Speech-Language Pathology, 10*(4), 323–342. [https://doi.org/10.1044/1058-0360\(2001\)028](https://doi.org/10.1044/1058-0360(2001)028)
- Eriks-Brophy, A., Quittenbaum, J., Anderson, D., & Nelson, T.** (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clinical Linguistics and Phonetics, 22*(8), 589–609. <https://doi.org/10.1080/02699200802221737>
- Evans, J. L., & Craig, H. K.** (1992). Language sample collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech, Language, and Hearing Research, 35*(2), 343–353. <https://doi.org/10.1044/jshr.3502.343>
- Finestack, L. H., Payesteh, B., Rentmeester Disher, J., & Julien, H. M.** (2014). Reporting child language sampling procedures. *Journal of Speech, Language, and Hearing Research, 57*(6), 2274–2279. https://doi.org/10.1044/2014_JSLHR-L-14-0093
- Fritz, C. O., Morris, P. E., & Richler, J. J.** (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18. <https://doi.org/10.1037/a0024338>
- George, S., Duran, N., & Norris, K.** (2014). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American Journal of Public Health, 104*(2), e16–e31. <https://doi.org/10.2105/AJPH.2013.301706>
- Guo, L.-Y., & Eisenberg, S.** (2015). Sample length affects the reliability of language sample measures in 3-year-olds: Evidence from parent-elicited conversational samples. *Language, Speech, and Hearing Services in Schools, 46*(2), 141–153. https://doi.org/10.1044/2015_LSHSS-14-0052
- Hadley, P. A.** (1998). Early verb-related vulnerability among children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1384–1397. <https://doi.org/10.1044/jslhr.4106.1384>
- Hall, S., Larrigan, L. B., & Madison, C. L.** (1991). A comparison of speech-language pathologists in rural and urban school districts in the state of Washington. *Language, Speech, and Hearing Services in Schools, 22*(4), 204–210. <https://doi.org/10.1044/0161-1461.2204.204>
- Hammer, C. S.** (2011). The importance of participant demographics. *American Journal of Speech-Language Pathology, 20*(4), 261–261. [https://doi.org/10.1044/1058-0360\(2011\)ed-04](https://doi.org/10.1044/1058-0360(2011)ed-04)
- Heilmann, J., Nockerts, A., & Miller, J. F.** (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393–404. [https://doi.org/10.1044/0161-1461\(2009\)09-0023](https://doi.org/10.1044/0161-1461(2009)09-0023)
- Hill, A., & Theodoros, D.** (2002). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare, 8*(4), 187–196. <https://doi.org/10.1258/135763302320272158>
- Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., Ganesalingam, K., Geier, P., Krass, C., Mathieson, S., McCabe, M., Micallef, K., Molomby, K., Ong, N., Pfeiffer, S., . . . Silove, N.** (2019). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *Journal of Telemedicine and Telecare, 25*(7), 431–437. <https://doi.org/10.1177/1357633X18776095>
- Hux, K., Morris-Friehe, M., & Sanger, D. D.** (1993). Language sampling practices: A survey of nine states. *Language, Speech, and Hearing Services in Schools, 24*(2), 84–91. <https://doi.org/10.1044/0161-1461.2402.84>
- Kramer, C. A., James, S. L., & Saxman, J. H.** (1979). A comparison of language samples elicited at home and in the clinic. *Journal of Speech and Hearing Disorders, 44*(3), 321–330. <https://doi.org/10.1044/jshd.4403.321>
- Larson, A. L., Barrett, T. S., & McConnell, S. R.** (2020). Exploring early childhood language environments: A comparison of language use, exposure, and interactions in the home and child-care settings. *Language, Speech, and Hearing Services in Schools, 51*(3), 706–719. https://doi.org/10.1044/2019_LSHSS-19-00066
- Miller, J. F., & Iglesias, A.** (2015). *Systematic Analysis of Language Transcripts* [Computer software]. Salt Software, LLC.
- Newkirk-Turner, B. L., Oetting, J. B., & Stockman, I. J.** (2016). Development of auxiliaries in young children learning African American English. *Language, Speech, and Hearing Services in Schools, 47*(3), 209–224. https://doi.org/10.1044/2016_LSHSS-15-0063
- Oetting, J. B., & McDonald, J. L.** (2002). Methods for characterizing participants' nonmainstream dialect use in child language research. *Journal of Speech, Language, and Hearing Research, 45*(3), 505–518. [https://doi.org/10.1044/1092-4388\(2002\)040](https://doi.org/10.1044/1092-4388(2002)040)
- Pezold, M. J., Imgrund, C. M., & Storkel, H. L.** (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools, 51*(1), 103–114. https://doi.org/10.1044/2019_LSHSS-18-0148
- Robertson, S. B.** (2019). *ASHA letter to ways & means rural and underserved communities health task force*. <https://www.asha.org/uploadedFiles/ASHA-Letter-to-WM-Rural-and-Underserved-112619.pdf>
- Ryan, C.** (2017). *Computer and Internet use in the United States: 2016*. U.S. Census Bureau.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition*. Pearson.
- Stockman, I. J.** (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27*(4), 355–366. <https://doi.org/10.1044/0161-1461.2704.355>
- Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J.** (2017). Telehealth language assessments using consumer grade equipment in rural and urban

-
- settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare*, 23(1), 106–115. <https://doi.org/10.1177/1357633X15623921>
- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L.** (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72(3), 748–767. <https://doi.org/10.1111/1467-8624.00313>
- Taylor, O. D., Armfield, N. R., Dodrill, P., & Smith, A. C.** (2014). A review of the efficacy and effectiveness of using telehealth for pediatric speech and language assessment. *Journal of Telemedicine and Telecare*, 20(7), 405–412. <https://doi.org/10.1177/1357633X14552388>
- U.S. Census Bureau.** (2018). *Poverty thresholds*. <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>
- Verdon, S., Wilson, L., Smith-Tamaray, M., & McAllister, L.** (2011). An investigation of equity of rural speech-language pathology services for children: A geographic perspective. *International Journal of Speech-Language Pathology*, 13(3), 239–250. <https://doi.org/10.3109/17549507.2011.573865>
- Waite, M. C., Cahill, L. M., Theodoros, D. G., Busuttin, S., & Russell, T. G.** (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare*, 12(S3), 92–94. <https://doi.org/10.1258/135763306779380048>
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M.** (2010). Internet-based telehealth assessment of language using the CELF-4. *Language, Speech, and Hearing Services in Schools*, 41(4), 445–458. [https://doi.org/10.1044/0161-1461\(2009/08-0131\)](https://doi.org/10.1044/0161-1461(2009/08-0131))
- Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M.** (2012). Assessing children's speech intelligibility and oral structures, and functions via an Internet-based telehealth system. *Journal of Telemedicine and Telecare*, 18(4), 198–203. <https://doi.org/10.1258/jtt.2012.111116>