Check for updates

Exploring the validity and reliability of online assessment for conversational, narrative, and expository discourse measures in school-aged children

Diana Burchell[1]*, Vincent Bourassa Bédard[2], Keara Boyce[3], Juliana McLaren[3], Myrto Brandeker[3], Bonita Squires[3], Elizabeth Kay-Raining Bird[3], Andrea MacLeod[4], Stefano Rezzonico[2], Xi Chen[1], Pat Cleave[3] and FrEnDS-CAN

[1]Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada, [2]École d'orthophonie et d'audiologie, University of Montréal, Montreal, QC, Canada, [3]School of Communication Sciences and Disorders, Dalhousie University, Halifax, NS, Canada, [4]Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

The COVID-19 pandemic has created novel challenges in the assessment of children's speech and language. Collecting valid data is crucial for researchers and clinicians, yet the evidence on how data collection procedures can validly be adapted to an online format is sparse. The urgent need for online assessments has highlighted possible the barriers such as testing reliability and validity that clinicians face during implementation. The present study describes the adapted procedures for on-line assessments and compares the outcomes for monolingual and bilingual children of online and in-person testing using conversational, narrative and expository discourse samples and a standardized vocabulary test. A sample of 127 (103 in-person, 24 online) English monolinguals and 78 (53 in-person, 25 online) simultaneous French-English bilinguals aged 7–12 years were studied. Discourse samples were analyzed for productivity, proficiency, and syntactic complexity. MANOVAs were used to compare on-line and in-person testing contexts and age in two monolingual and bilingual school-age children. No differences across testing contexts were found for receptive vocabulary or narrative discourse. However, some modality differences were found for conversational and expository. The results from the study contribute to understanding how clinical assessment can be adapted for online format in school-aged children.

KEYWORDS

discourse measures, speech-language pathology, online assessment, children, conversation, expository, narrative

Introduction

During the COVID-19 pandemic, practitioners have been searching for feasible adaptations to current language assessment practices, since traditional in-person assessments have not been possible. Being able to collect valid data is crucial for researchers and clinicians, yet the evidence on how data collection procedures specific to language samples can validly be adapted to an online format is sparse (Taylor et al., 2014; d'Orville, 2020; Reimers et al., 2020). Online assessments of language, traditionally used for remote clients, have now been widely implemented by researchers and clinicians alike due to the sudden lockdown of in-person services in the beginning of 2020 (e.g., Mansuri et al., 2021). Emerging evidence is looking at the developments necessary to make tele-practice a reliable and valid assessment method (Chenneville and Schwartz-Mette, 2020; Putri et al., 2020; Fong et al., 2021). The present study will describe the adapted procedures required to pursue a large-scale study of discourse in typically developing monolingual and bilingual school-aged children. We will examine whether online and in-person conversation, narrative or expository discourse sample measures or standardized vocabulary tests differ for monolingual and bilingual children, and, if so, what might account for these differences.

Few studies have investigated the online data collection of discourse samples among monolingual and bilingual school-aged children. There are a few studies on monolingual adults in clinical populations (e.g., Turkstra et al., 2012), as well as studies of both monolingual (Manzanares and Kan, 2014) and bilingual (Guiberson et al., 2015) preschool children. One systematic review by Taylor et al. (2014) evaluated the efficacy and effectiveness of speech and language assessments online. The authors found 5 studies who met the inclusion criteria but stated that the articles were of variable quality and did not provide enough evidence to influence clinical practice. This review presented evidence of inter-rater reliability in online language assessments, and most studies found sufficient inter-rater reliability to suggest the contexts did not significantly alter the results. They confirmed that more rigorous statistics are needed to either confirm or dispute this preliminary evidence. More recently, Manning et al. (2020) looked at the feasibility, reliability, and validity of obtaining language samples remotely by recording child-parent play with toddlers. They compared online and in-person groups on language sample metrics such as mean length of utterance (MLU), number of different words (NDW), and type-token ratio (TTR). This study found no evidence of differences across any language metrics due to modality. To the best of our knowledge, there are few published studies that have investigated the comparability of online and in-person assessments of school-age children using rigorous and parametric statistics. The following paragraphs will provide an overview of research relating to measures of vocabulary, and

more in-depth review of measures of discourse and language metrics. While the larger project included both macrostructural (i.e., related to the meaning conveyed) and microstructural (i.e., related to the language used) measures, for the purposes of this paper, only microstructural measures were included. This was a strategic decision since microstructural measures represent language development, which is of particular interest to clinicians during the pandemic.

Vocabulary measures have been previously validated for use in online assessment. Haaf et al. (1999) created and evaluated two online measures of the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn and Dunn, 1981). They found that the online versions of the PPVT-R were not significantly different from the in-person version of the PPVT-R. The authors concluded that an online version of the PPVT is statistically equivalent to the in-person version and can be used in conjunction with the published norms. Eriks-Brophy et al. (2008) later confirmed this statistical equivalence using an updated version of the PPVT (PPVT-III, Dunn and Dunn, 1997). This replication of Haaf's original study with a new version of the PPVT indicates the stability of these findings, even with slight methodological changes.

In general terms, discourse is typically defined as a conversation between people as a form of communication. Within the fields of linguistics and speech language pathology, discourse is more specifically defined as 'a linguistic unit (such as conversation or a story) larger than a sentence' [Merriam-Webster, (n.d)]. Discourse skills have been shown to be critical to school success and are known to be an area of difficulty for children with language-learning disabilities (Paul and Norbury, 2012). Three main types of discourse include conversation, exposition, and narration. Conversation has been defined as a "dialogue between people where each contributes by making statements, asking questions, and responding to the other speaker" (Nippold et al., 2014, p. 877). Exposition and narration are monologic in clinical settings, where expository discourse is defined as the use of language to convey information (Bliss, 2002) while narrative discourse is defined as telling stories about oneself and/or others (Nippold et al., 2014).

Children develop discourse skills over a long period of time and across a variety of genres. They begin to develop the ability to engage in conversational discourse even before they start to speak, and this skill continues to be refined through the school years (Hoff, 2009). Expository discourse begins to develop later than conversation and narration. Procedural description, persuasion, negotiation and explanation are all forms of expository discourse (Nippold et al., 2007; Nippold and Sun, 2010). Expository discourse emerges within conversations in the preschool period (Cabell et al., 2011) but becomes more prevalent in children's experiences once schooling begins and is increasingly more frequent in their spoken language at that time (Nippold and Sun, 2010). In narration, children begin to

talk about past events and to produce brief narrative recounts of these events by the age of two when scaffolded by a parent (Eisenberg, 1985). By 5 years of age, they are able to produce narratives with some plot structure (Hoff, 2009; Owens, 2012) and the complexity of their spoken narratives continues to develop through at least 12 years of age (Hoff, 2009; Cabell et al., 2011). The following paragraphs will provide an overview of the literature investigating and comparing language use in these three discourse genres.

Conversational tasks have been the most effective in accurately portraying the discourse level skills of younger children (Leadholm and Miller, 1992; Heilmann et al., 2010). Furthermore, conversational tasks are more reflective of basic interpersonal communication skills (e.g., BICS) as opposed to later developing discourse tasks that may be more in line with curriculum expectations and cognitive academic language proficiency (e.g., CALP), such as expository and narrative measures (Heilmann et al., 2010). However, assessing basic interpersonal language in school-age children may still be useful for clinicians, especially in the case of language learners who may not have developed adequate academic language yet (see Cummins, 2000 for a review of BICS and CALP).

Expository tasks are highly structured measures which focus on explaining a specific topic (Berman and Nir-Sagiv, 2007). Furthermore, expository tasks can be curriculum based, which can be a powerful diagnostic tool in evaluating children's expressive language (Heilmann and Malone, 2014). There is preliminary evidence that expository tasks accurately capture the development of academic language skills (Kay-Raining Bird et al., 2016). This is supported by a study conducted by Nippold et al. (2005), which compared conversational and expository discourse. The authors found that students demonstrated greater syntactic complexity on the expository task, indicating that complex thought underlies complex language. In the present study, looking at an expository measure is useful as an index of both academic language and curriculum-based assessment.

Researchers such as Stadler and Ward (2005) have shown that narrative skills are a rich reflection of children's oral language development. This is because narratives require more complex vocabulary and an overarching structure. Narrative skills are typically assessed in two ways, wherein students are asked to demonstrate comprehension of a "model" story and/or produce an original story. Storytelling skills emerge in the preschool period and continue to grow throughout their time in school. Narrative development is supported through activities such as storybook reading. As students' oral language competency grows, it increases their complexity of their language (Verhoeven and Strömqvist, 2001). In this study, on-line and in-person narrative production tasks were compared.

Discourse tasks provide a context for observing children's language abilities, and specific language metrics allow researchers and clinicians to obtain quantitative observations

that can reflect children's development and proficiency. For the purposes of this paper, we chose to include the most commonly used language metrics, which includes measures of productivity, proficiency, and syntactic complexity (Schneider et al., 2004). Lexical productivity is a metric which measures the amount of output generated by a participant (Le Normand et al., 2008). Studies have shown that lexical productivity tends to increase with age for both monolingual and bilingual children (Le Normand et al., 2008; Jia et al., 2014). Productivity was of particular interest in this study since it has been correlated with psycho-social variables such as introversion, anxiety and shyness (see Dewaele and Pavlenko, 2003 for a review). Emerging research during the COVID-19 pandemic has shown increased levels of anxiety and shyness due to online schooling and social isolation (Imran et al., 2020; Lavigne-Cerván et al., 2021; Orgilés et al., 2021). It is therefore vital to ascertain whether psycho-social factors such as anxiety and shyness may affect the administration of online discourse-level skills for clinicians.

Multiple studies attest to the critical role of syntactic complexity in the development of language and literacy skills in school-age populations. However, syntactic complexity is dependent on the type of task administered: studies show that children produce more complex utterances during expository discourse than they do in conversation (e.g., Nippold, 2009). There is also emerging evidence that modeling (used in our narrative assessments), which involves syntactic priming, may also impact syntactic complexity (Zebib et al., 2020). Syntactic complexity is therefore an interesting metric in this study, since it can be variable across tasks but may be stable across modality contexts.

Mean length of C-Unit in morphemes (MLCUm) is a micro-structural (i.e., linguistic) measure that is a general reflection of both general language proficiency as well as the syntactic complexity of the discourse being analyzed (Craig et al., 1998; Eisenberg et al., 2001). An utterance is defined as one main clause and all dependent clauses associated with it (Miller et al., 2006). With increased language proficiency, children will begin to incorporate more advanced linguistic devices into their speech, including conjunctions and subordinate clauses, resulting in greater MLCUm values (Berman and Slobin, 1994). However, in the past, researchers have found that MLCUm changes with age, improving more significantly in expository and narrative discourse than in conversation for older students (i.e., teen years) (Leadholm and Miller, 1992; Rice et al., 2010; Westerveld and Moran, 2013). In this study, MLCUm is one metric used to measure whether children are showing the same proficiency on-line as they would in person. Similar to syntactic complexity, we expect that MLCUm may be stable across modality contexts.

The current study emerged from the project "*French/English Discourse Study – Canada*" (FrEnDS-CAN), which focuses on a variety of discourse skills in typically developing

TABLE 1 Frequency of gender and age group by condition and language group.

		Monolingual		Bilingual	
		In-person ($n = 103$)	Online ($n = 24$)	In-person ($n = 53$)	Online ($n = 25$)
Gender	Male	34	10	22	2
	Female	44	13	19	5
	Not disclosed	25	1	12	18
Age group	7–8 years	48	5	19	10
	9–10 years	38	10	18	10
	11–12 years	17	9	16	5

monolingual and bilingual school-aged children. Typically developing children were chosen as a population of study since they constitute a first step in better understanding what is expected in school-aged children and may serve as foundation for future research studies of children with language or learning disorders. The project is set in five Canadian cities (Halifax, Moncton, Montréal, Ottawa, and Toronto) and data collection started with in-person procedures in 2016. In March 2020, an online data collection procedure was adapted. The present study describes the adapted procedures and compares the outcomes of online and in-person testing using discourse samples and standardized vocabulary testing for monolingual and bilingual children. The present study also examines whether the impact of modality differs across measures and what might account for these differences. Specifically, we asked:

1. Did discourse (conversation, expository, narration) or standardized vocabulary measures differ when testing was done in-person vs. online?
2. Did the impact of modality vary across productivity, proficiency, and syntactic complexity?
3. Did the impact of modality vary with age from 7 to 12 years of age?

Materials and methods

Participants

This study used a subset of data collected for a larger Canadian French/English Discourse study (FrEnDS-CAN) investigation of discourse development in school-age bilingual and monolingual children. For the present analyses, 127 (103 in-person, 24 online) English monolinguals and 78 (53 in-person, 25 online) simultaneous French-English bilinguals were included. The children were distributed across three age groups (7–8, 9–10, and 11–12 years). For the monolingual children, there were 57 female children (44 male and 26 who did not disclose) and the mean age was 8.57 (SD = 1.525). For the simultaneous bilingual children, there were 24 female children

(24 male and 30 who did not disclose) and the mean age was 8.79 (SD = 1.598). For more information about how our participants were distributed across groups, please see Table 1 below. All children were typically developing, with no diagnosed or suspected language, hearing or learning difficulties (as established through parent report). They were recruited through their schools, through posters placed in public places and on social media.

The monolingual children were recruited through research teams in New Brunswick, Nova Scotia, Ontario and Quebec. They were exposed to only English in the home and attended English-language schools. Since Canada has two official languages, English-schooled children are required to take French as an academic subject (core French) in elementary school starting in grade 4 in both provinces. Thus, children were considered English monolinguals if they were exposed to French <10% of the time (e.g., only through these core classes at school), established through parent report.

The simultaneous bilingual children were recruited through research teams in Montréal (Quebec), Ottawa (Ontario), and Moncton (New Brunswick). They were exposed to both English and French from before the age of three and were able to complete testing in each language. In the home language questionnaire, there were 18 students who primarily spoke English at home, 19 who primarily spoke French at home, and 9 who reported an even split between the two languages. These children attended French-language schools and lived in communities where both French and English were spoken, so were likely to encounter both languages outside the home. In French-language schools, English classes are required starting in grades 1 or 2 in Quebec, grade 4 in Ontario, and grade 3 in New Brunswick. While these children were assessed in both languages, for the purposes of this study, we will only be discussing their English performance.

Procedures

Ethical approval was obtained through each participating university and school district. Consent forms were distributed

and collected through schools or *via* email. Parents confirmed their child's eligibility to participate by checking appropriate boxes on the consent form. Monolinguals were tested in a single session; bilinguals were tested in two sessions by different examiners: one in English, the other in French. Parallel tests and tasks were administered in French and English. The order of language testing was counterbalanced such that equal numbers of children within every age group were tested in French or English first. In each language, the test protocol began with a standardized vocabulary comprehension test. This was followed by conversational, narrative and expository language samples, in which the order of administration was also counterbalanced within age groups and within each language for bilinguals. The in-person, but not the online protocol, ended with the administration of a non-word repetition task. Since the non-word repetition task was not administered in both contexts, it is not discussed further. Any French language testing is also not included in this study and will not be discussed further. Examiners were graduate students, undergraduate students, or researchers. All student examiners were trained on test and task administration by the same Ph.D. student.

In-person testing occurred in a quiet area of the child's school or in a testing room in the research laboratories of participating universities. For in-person testing, sessions were recorded using a digital voice recorder which was placed next to the child. In March 2020, in-person testing was suspended in response to the COVID-19 pandemic. Consequently, materials were modified to accommodate online testing (discussed in the materials section). Zoom and Microsoft TEAMS platforms were used to test children online. The research team trained the testers in the use of the online platforms and the online administration of the full testing protocol was piloted on two children (9 and 10 years of age). These videotaped pilot sessions were then used as examples to train other testers. Children who were tested online used their home computer in a quiet area of their house. For online testing, sessions were recorded locally on the tester's computer using the platform of the parent's choice (either Zoom or Microsoft Teams). When technical issues arose which impeded the audio quality, testers would stop testing and troubleshoot the connection with the family until the audio quality was sufficient. A parent was asked to be available during testing in case the child experienced any technical or other difficulties. This usually meant they were in the room with the child but not sitting with them. Testers made notes of any significant behavioral or technical issues that arose during online testing using a common form.

Materials

Vocabulary comprehension

The *Peabody Picture Vocabulary Test-4* (*PPVT-4*; Dunn and Dunn, 2007) was used to test vocabulary comprehension in

English. The *PPVT-4* has good reliability and validity and is commonly used for both educational and research purposes. In this vocabulary comprehension test, children point to one of four pictures in response to spoken word stimuli. Each test has a start point determined by age. Basals and ceilings were determined following manual instructions. In-person testing used the test booklet. Online testing followed the same procedures as in-person testing except that the stimuli were presented as images and the examiner's screen was shared with the child; each image presented a different item's picture stimuli.

Conversational samples

A conversational sample of at least 10 min was collected by the adult examiner following the *Systematic Analysis of Language Transcripts* (*SALT*; Miller and Chapman, 2012) interview protocol. This involved talking to the child about topics of interest to them. The child was initially asked what they would like to talk about and as the conversation progressed, additional topics from a common set (e.g., family, pets, school, hobbies) were introduced if needed. This flexible protocol was selected since it allowed children to choose a topic they were motivated to speak about, and therefore gave them an opportunity to demonstrate their oral language abilities. All examiners were instructed to listen as often as possible, and to only participate in the conversation when necessary to keep the child engaged. For example, examiners were instructed to ask open-ended questions such as "What were your favorite memories from your trip?" or "What do you like about art class?"

Narrative samples

Two narrative samples were collected in each language, the first using a story stem (setting information provided to generate a story) and the second using a single picture elicitation task. The story stem narratives are not analyzed here, so will not be discussed further. The single-picture elicitation was one of three narrative tasks included in the *Test of Narrative Language-2* (*TNL-2*; Gillam and Pearson, 2017; the revised test materials were shared with the research team prior to publication). It uses a 'give a story, get a story' format in which the examiner first tells a story about a complex picture (i.e., two children hiding behind a rock watching a treasure chest with either a dragon or pirates guarding it), asks 12 comprehension questions (6 literal, 6 inferential) about the treasure story, then produces a second picture (two children hiding as a family of aliens deboards from a spaceship or two children watching as an ogre has a Pegasus on a rope) and asks the child to produce "an even better" story using this new picture. Throughout this task, the child and the researcher were always able to see and reference the pictures (e.g., the pictures were either placed on a desk before the child or the researcher shared their screen with the pictures on it). The *TNL-2* is widely used and has high reliability and validity.

Expository samples

Expository samples were collected using an adaptation of "The Favorite Game or Sport Task" developed by Nippold et al. (2005) and modified by Heilmann and Malone (2014). In this protocol, children were asked to describe how to play a game or sport of their choice. Games without clear rules or an ending and videogames were excluded. Children first identified the game or sport they would describe. They were given a few minutes to plan what they wished to say. To encourage them to think broadly, the children were presented with eight areas they might discuss: What you try to do, Getting ready to play, Starting the game, How you play, Rules, Scoring, and Ending the game. During in-person testing, each of these components was printed on a two- by four-inch card and presented randomly in an array in front of the child with a brief verbal explanation of each (e.g., "you could talk about what you're trying to do in the game"). If asked, the examiner could re-read the cards; this occurred occasionally with younger children. When planning, the children could rearrange the cards as needed although only a minority of children did so. For online testing, circles with these same components printed in them were presented individually *via* screen sharing, until all were present on the screen where they remained for the duration of the task. One of ten pre-determined orders of presentation were used, selected randomly by the examiner. Once all the topic areas were introduced either on the computer screen or with cards, the child was given as much time as they wished to plan. When they said they were ready, they were asked to explain the game. When they indicated they were done, the examiner asked the child to explain any special strategies that could be used to win the game.

Analyses

Transcription

Each discourse sample was transcribed into Communication-units (C-units) by trained graduate students. A C-unit is defined as an independent clause and its modifiers (Hughes et al., 1997); it may be incomplete (e.g., a few words in response to a question). Transcriptions followed slightly modified *SALT* conventions. The modifications included: writing contractions as separate words rather than slashing them (e.g., don't = do not) and slashing the past participle –en (e.g., was give/en). Lexical verbs were identified using a [v] code next to the verb (e.g., was give/en[v]). Transcription of conversational samples began as soon as the child was interacting naturally and continued for 10 consecutive minutes. If there was not 10 min of conversation available, additional time was taken from conversations between child and examiner throughout the session to obtain the full 10 min. Narrative and expository transcriptions began after the instructions were completed and ended when the child indicated they were finished. Sample transcripts were saved in separate files and checked and corrected with reference to the session's audio-recording by a second, experienced transcriber.

Microstructure metrics

SALT software was used to generate microstructure metrics separately for each discourse file. For the purposes of this study, three microstructure metrics were computed. The first was a language proficiency metric, the mean length of utterance in morphemes (MLCUm). This is automatically calculated in *SALT* by averaging the number of morphemes (i.e., word roots and slashed morphemes) per C-unit. This study uses morphemes instead of words since that metric is more commonly utilized by clinicians and since this is not a cross-linguistics study. Secondly, we looked at a productivity metric, the number of total words produced by the child, referred to as NTW. Finally, we looked at syntactic complexity. In this case, we created a syntactic complexity score (SC) in SPSS by dividing the number of lexical verbs by the total number of C-units. These metrics were specifically chosen because they represent the three microstructure measures that might interest clinicians in online assessment: language proficiency, productivity, and syntactic complexity.

Design

To investigate the impact of modality on language assessment, we tested two groups of school-aged children in English: monolingual and simultaneous bilingual. Three age groups were included: 7–8-, 9–10-, and 11–12-year-olds. Three different language samples were collected from each participant: conversation, expository and narrative. Within each of these samples, we looked at three micro-structure measures (dependent variables): mean length of C-Unit in morphemes (MLCUm), number of total words (NTW) and syntactic complexity (SC). We also analyzed the raw scores of a standardized test of vocabulary (PPVT). The participants were tested either online or in-person.

Statistical analyses

All statistical analyses were run in SPSS version 28. Descriptive statistics (means, standard deviations) were generated on all English standardized test scores and discourse sample measures, separately for children tested in-person and online. Descriptive statistics were computed for the whole group of monolinguals and for the whole group of bilinguals as well as for each of the three age groups within those groups. We then completed two types of statistical analyses: ANOVAs for vocabulary analyses or MANOVAs for discourse analyses, and Bayesian t-tests.

A two-way modality (online vs. in-person) by age (7–8, 9–10, 11–12) between-subjects ANOVA tested mean differences

on raw PPVT scores, separately for monolingual and bilingual groups. Two-way modality (online vs. in-person) by age (7–8, 9–10, 11–12) between-subjects MANOVAs tested mean differences in the three discourse measures, separately for conversation, exposition, and narration for the two language groups (monolingual and bilingual). Preliminary analyses for the MANOVAs were completed. Significant main effects for age were examined with *post-hoc* comparisons with a Bonferroni correction for alpha level. Boxplots showed there were no outliers. The data were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > 0.05$) and there was homogeneity of variances ($p > 0.05$) and covariances ($p > 0.05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Alpha was set at 0.05 a priori for all analyses.

Bayesian statistics were used to follow-up when non-significant modality main effects were obtained. The Bayes factor (BF_{01}) statistic was used. Bayes factors globally confirm that the absence of difference is not due to a lack of power but to the fact that the two modalities are equal (Brydges and Gaeta, 2019). A $BF_{01} > 1$ indicates evidence for the null hypothesis (H0). The further a value is from 1 (up to 100) the stronger the evidence is in favor of the null hypothesis (IBM, 2021; van Doorn et al., 2021).

Results

Monolinguals

Table 2 provides the means and standard deviations for PPVT scores and the conversation, expository, and narrative discourse measures. For a summary of the results, please see Table 1 in Appendix 1.

Vocabulary

The two-way ANOVA was conducted on receptive vocabulary raw scores. Only the main effect for age was significant, $F_{(2,105)} = 11.808$, $p < 0.001$, partial $\eta^2 = 0.191$. *Post-hoc* paired comparisons on the age effect found vocabulary raw scores differed significantly for all age groups and increased with increasing age (7–8: M = 141.57, SD = 16.70; 9–10: M = 157.92, SD = 18.31; 11–12, M = 176.26, SD = 16.74). The follow-up Bayesian analysis found a BF_{01} of 2.106, which provided only anecdotal evidence for the equivalence of a modality effect for raw receptive vocabulary scores in these monolingual children.

Conversational discourse

A two-way MANOVA analyzed MLCUm, NTW and SC in conversation. No significant main effects or interactions were obtained for MLCUm or SC. However, for NTW, the

main effects of modality [$F_{(1,120)} = 5.579$, $p = 0.020$, partial $\eta^2 = 0.046$] and age [$F_{(2,120)} = 4.158$, $p = 0.018$, partial $\eta^2 = 0.067$] were significant, but not the interaction. In terms of modality, more words were produced in-person (M = 839.98; SD = 239.21) than on-line (M = 759.64; SD = 278.12) in conversations. *Post-hoc* paired comparisons showed the number of words produced in conversation increased significantly from 7 to 8 years (M = 721.24, SD = 228.98) to both 9–10 (M = 880.98, SD = 222.40) and 11–12 (M = 923.23, SD = 259.47) years of age. Bayesian statistics confirmed moderate evidence that the two modalities were equivalent for MLCUm ($BF_{01} = 5.14$) and SC ($BF_{01} = 5.51$).

Expository discourse

A two-way MANOVA analyzed MLCUm, NTW and SC in expository samples. For MLCUm, significant main effects were obtained for modality [$F_{(1,120)} = 5.406$, $p = 0.022$, partial $\eta^2 = 0.045$] and age [$F_{(2,120)} = 3.196$, $p = 0.045$, partial $\eta^2 = 0.053$]. Additionally, the main effect of age was significant for NTW [$F_{(2,120)} = 3.421$, $p = 0.036$, partial $\eta^2 = 0.056$]. No other main effects or interactions reached significance. The modality effect indicated greater MLCUm for in-person (M = 11.30, SD = 2.20) compared to on-line (M = 10.48, SD = 1.48) expository samples. *Post-hoc* paired comparisons of age groups indicated that the 11–12 group (M = 12.08, SD = 1.69) produced significantly longer C-units in expository samples than the 7–8 group (M = 10.49, SD = 2.33). Additionally, the 11–12 group produced significantly more words (M = 620.04, SD = 328.75) than the 7–8 age group (M = 378.02, SD = 247.58) in these samples.

Follow-up Bayesian analyses showed Bayesian factors of $BF_{01} = 3.59$ for NTW and $BF_{01} = 2.96$ for SC. These indicated the strength of evidence was moderate for NTW but anecdotal for SC that the two modalities were equivalent.

Narrative discourse

The narrative MANOVA revealed a main effect of age for SC [$F_{(2,120)} = 10.244$, $p < 0.001$, partial $\eta^2 = 0.153$]. No other main effects or interactions were obtained. *Post-hoc* paired comparisons of the age effect revealed that the youngest group (M = 1.28, SD = 0.35) produced significantly fewer verbs per C-unit than either the middle (M = 1.57, SD = 0.37) or oldest (M = 1.63, SD = 0.32) age groups.

Bayesian follow-up analyses showed a $BF_{01} = 1.242$ for MLCUm, $BF_{01} = 3.337$ for NTW, and $BF_{01} = 2.375$ for SC. These scores provide moderate evidence of modality equivalence for NTW and anecdotal evidence for the other narrative discourse measures.

TABLE 2 Means and standard deviations of monolingual participants on all measures.

		7–8 years ($n = 53$)		9–10 years ($n = 48$)		11–12 years ($n = 26$)	
		In-person ($n = 48$)	Online ($n = 5$)	In-person ($n = 38$)	Online ($n = 10$)	In-person ($n = 17$)	Online ($n = 9$)
Receptive vocabulary (PPVT raw score)							
	Mean	140.8	149.0	157.09	160.9	177.9	171.7
	SD	17.03	12.27	18.43	18.64	16.03	19.41
Conversation							
MLCUm	Mean	7.37	7.783	7.55	7.368	7.67	7.85
	SD	1.71	2.719	1.24	1.11	1.24	1.65
NTW	Mean	729.4	629.5	909.8	762.7	980.8	814.4
	SD	233.4	307.3	206.5	258.5	224.8	298.2
SC	Mean	0.905	1.00	0.924	0.815	0.926	0.956
	SD	0.231	0.357	0.177	0.138	0.184	0.282
Expository							
MLCUm	Mean	10.50	10.37	11.68	9.89	12.56	11.18
	SD	2.41	1.64	1.72	1.53	1.76	1.17
NTW	Mean	371.8	432.4	395.9	492.9	664.1	536.9
	SD	254.8	183.6	231.6	299.3	355.9	269.1
SC	Mean	1.82	1.43	1.73	1.39	1.71	1.45
	SD	2.02	0.36	0.384	0.308	0.387	0.264
Narrative							
MLCUm	Mean	8.71	9.1	9.46	10.33	9.94	10.68
	SD	2.37	1.34	2.77	1.72	1.30	2.91
NTW	Mean	203.5	236.8	246.4	276.9	347.8	339.6
	SD	173.6	265.5	149.9	163.2	248	184
SC	Mean	1.30	1.08	1.53	1.69	1.60	1.68
	SD	0.342	0.439	0.377	0.337	0.258	0.447

PPVT, Peabody Picture Vocabulary Test, 4th edition (Dunn and Dunn, 2007); MLCUm, mean length of communication unit in morphemes; NTW, number of total words; SC, syntax complexity; SD, standard deviation.

Simultaneous bilinguals

Only the English tasks were analyzed for the bilingual children to be congruent with the monolingual group. Table 3 shows the means and standard deviations for vocabulary (PPVT) scores and conversation, expository, and narrative discourse measures. For a summary of the results, please see Table 1 in Appendix 1.

Vocabulary

The two-way ANOVA revealed only a significant main effect of age for these bilingual children, $F_{(2,53)} = 18.52$, $p < 0.001$, partial $\eta^2 = 0.421$. *Post-hoc* paired comparisons showed that PPVT raw scores were significantly lower for the 7–8-year-olds (M = 124.8, SD = 24.95) than either the 9–10 (M = 157.70, SD = 26.85) or 11–12-year-olds (M = 176.07, SD = 14.82).

Bayesian follow-up analysis revealed a BF_{01} of 4.499. This provided moderate evidence for modality equivalence on receptive vocabulary.

Conversational discourse

The two-way MANOVA analyzing conversational discourse measures revealed significant main effects of age for MLCUm [$F_{(2,76)} = 6.156$, $p = 0.003$, partial $\eta^2 = 0.148$], NTW [$F_{(2,76)} = 7.156$, $p < 0.001$, partial $\eta^2 = 0.168$], and SC [$F_{(2,76)} = 3.861$, $p = 0.026$, partial $\eta^2 = 0.098$]. No other main effects or interactions were significant. All three measures increased with age. *Post-hoc* paired comparisons revealed significant improvement in MLCUm from 7 to 8 (M = 6.69, SD = 1.54) to 11–12 (M = 8.25, SD = 1.62). A similar pattern was found for the NTW (7–8: M = 635.07, SD = 257.94; 11–12: M = 907.43, SD = 256.23) and SC (7–8: M = 0.79, SD = 0.24; 11–12: M = 0.98, SD = 0.26) measures.

TABLE 3 Means and standard deviations of bilingual participants on all measures.

			7–8 years ($n = 29$)		9–10 years ($n = 28$)		11–12 years ($n = 21$)	
			In-person ($n = 19$)	Online ($n = 10$)	In-person ($n = 18$)	Online ($n = 10$)	In-person ($n = 16$)	Online ($n = 5$)
Receptive vocabulary (PPVT raw score)								
		Mean	125.5	123.4	152.6	165.6	174.2	183.0
		SD	27.26	21.9	26.8	26.45	13.33	21.17
Conversation								
	MLCUm	Mean	6.807	6.480	7.503	7.866	8.081	8.772
		SD	1.462	1.726	1.880	1.307	1.572	1.835
	NTW	Mean	672.8	563.2	702.6	840.1	904.1	918.2
		SD	273.3	221.1	240.2	152.4	247.6	313.4
	SC	Mean	0.807	0.751	0.831	0.875	0.976	1.001
		SD	0.255	0.212	0.243	0.114	0.239	0.337
Expository								
	MLCUm	Mean	9.774	9.671	11.16	10.21	11.87	11.28
		SD	2.273	1.398	4.428	1.648	2.106	3.626
	NTW	Mean	303.4	330.7	504.5	507.5	560.8	675.6
		SD	223.9	285.5	565.2	317.1	283.4	301.9
	SC	Mean	1.477	1.357	1.494	1.414	1.707	1.526
		SD	0.515	0.420	0.378	0.333	0.407	0.567
Narrative								
	MLCUm	Mean	8.213	8.708	9.562	10.46	9.744	10.75
		SD	1.450	1.756	2.360	1.958	1.380	2.142
	NTW	Mean	243.1	134.7	208.9	232.4	327.4	340.6
		SD	152.0	87.72	120.1	61.44	213.3	149.4
	SC	Mean	1.226	1.207	1.385	1.621	1.554	1.799
		SD	0.262	0.302	0.436	0.376	0.208	0.551

PPVT, Peabody Picture Vocabulary Test, 4th edition (Dunn and Dunn, 2007); MLCUm, mean length of communication unit in morphemes; NTW, number of total words; SC, syntax complexity; SD, standard deviation.

Bayesian results were: $BF_{01} = 5.340$ for MLCUm, $BF_{01} = 5.295$ for NTW, and $BF_{01} = 5.209$ for SC. Thus, moderate evidence supports the conclusion that the two modalities were equivalent for all of the conversational measures for the bilingual group.

Expository discourse

Two-way MANOVA results showed a main effect of age for MLCUm [$F_{(2,76)} = 3.513$, $p = 0.035$, partial $\eta^2 = 0.091$] and NTW [$F_{(2,76)} = 3.807$, $p = 0.027$, partial $\eta^2 = 0.098$] only. No other main effects or interactions were significant. *Post-hoc* paired comparisons showed lower MLCUm scores for 7–8 (M = 9.74, SD = 1.99) than 11–12 (M = 11.73, SD = 2.45) age groups as they were for NTW scores (7–8: M = 312.83, SD = 242.10; 11–12: M = 588.14, SD = 284.57).

Bayesian analyses resulted in a MLCUm BF_{01} of 4.43, a NTW BF_{01} of 5.12, and a SC BF_{01} of 2.42. Leading us to conclude there is moderate evidence (anecdotal

for SC) for modality equivalence between groups on any expository measure.

Narrative discourse

Finally, a two-way MANOVA revealed main effects of age for all narrative measures: MLCUm [$F_{(2,76)} = 6.295$, $p = 0.003$, partial $\eta^2 = 0.152$], NTW [$F_{(2,76)} = 4.939$, $p = 0.010$, partial $\eta^2 = 0.124$], and SC [$F_{(2,76)} = 9.338$, $p < 0.001$, partial $\eta^2 = 0.211$]. No other main effects and interactions were significant. *Post-hoc* paired comparisons showed 7–8 MLCUm (M = 8.39, SD = 1.55) to be significantly lower than either 9–10 (M = 9.88, SD = 2.23) or 11–12 (M = 10.00, SD = 1.60) age groups. This was also true for SC (7–8: M = 1.22, SD = 0.27, 9–10: M = 1.47, SD = 0.42; 11–12: M = 1.62, SD = 0.33). In contrast, 7–8-year-olds (M = 204.39, SD = 140.96) produced fewer words (NTW) in narratives than 11–12-year-olds only (M = 330.70, SD = 195.55); and the Bayesian revealed anecdotal certainty that there were no significant effects of modality for

MLCUm ($BF_{01} = 0.8720$), and SC ($BF_01 = 0.1126$). However, for NTW, the BF_{01} was 3.577, indicating moderate evidence of modality equivalence.

Discussion

The current study examined the comparability of online and in-person assessment on conversational, expository, and narrative discourse across both monolingual and simultaneous bilingual speakers of English. Specifically, we looked at metrics of productivity, proficiency, and syntactic complexity across these three forms of discourse. We furthermore examined the effect of age against the previous two questions. Overall, our results indicated that most measures seem to be comparable across in-person and online assessment contexts. For the monolingual group, there were no differences due to modality on either vocabulary or narrative measures. However, there were two distinct differences due to modality for conversation. First, we saw an impact of modality on the productivity metric of the conversational measure in favor of the in-person group. Second, we saw an impact of modality on the proficiency and syntactic complexity measure of the expository measure in favor of the in-person group. Finally, while students improved with age, the effect of modality did not vary with the age of the participants.

For the simultaneous bilingual group, we saw no differences across vocabulary, conversation, expository or narrative. More specifically, we saw no differences in either productivity or syntactic complexity due to the assessment context. While we did see that students improved with age on these measures, age did not have an impact on the effect of assessment modality. It is interesting to note that, despite differing in exposure to and use of the target language compared to their monolingual counterparts, we did not see differences due to assessment modality for the simultaneous bilingual group. However, it should also be noted that the simultaneous bilingual group was smaller, which may have affected the power level of these analyses. The theoretical and clinical implications of these findings will be discussed below.

Across both language groups and all ages, we saw no differences on receptive vocabulary when comparing the in-person and online assessment groups. We speculate that this can be explained by several factors. First, and most importantly, the PPVT had already been adapted to online assessment and validated. It was already common practice to use an electronic version t of the PPVT instead of a paper copy, even in person, which was very simple to use in an online format. Furthermore, the PPVT can be more easily adapted for online assessment difficulties. For example, if children are shy or anxious about the session, they can either hold up the number of fingers to indicate which picture they choose, or they can type it in the chat. There is also less impact due to Wi-Fi or audio issues since both researchers and participants are only communicating single

words (e.g., "Picking" from examiner and "two" from the child). We can therefore state with some degree of confidence that the results of receptive vocabulary tasks are comparable across in-person and online testing contexts.

For our conversational measure, we saw no differences across either language group on syntactic complexity. There were also no differences in modality due to the age group of the participant. However, in the monolingual group, we did see a significant difference between the online and in-person testing groups, in favor of the in-person group, on the number of total words (NTW), which is a measure of overall productivity. This may be attributed to several things. First, the context itself caused changes to the task that the researchers could not control. For example, the Wi-Fi and audio quality seemed to impact the conversational measure the most. Researchers often had to ask students to repeat themselves, which can cause students to withdraw. If students froze, they would lose their train of thought and have more difficulty getting their momentum back. It also seemed like students had more difficulty engaging with the task if they experienced interruptions. Furthermore, the conversational measure is the least structured of all the tasks. We suspect that the conversational measure shows differences between contexts on productivity because it is the only turn-taking task and is therefore more susceptible to factors such as lags in audio quality, lack of gestural language, less fluidity, and less non-verbal cues. For shy or anxious students, even in person, this can be a daunting task. Emerging research has shown that the current COVID-19 pandemic has exacerbated these challenges such as language anxiety (Imran et al., 2020; Lavigne-Cerván et al., 2021; Orgilés et al., 2021). In the clinical implications section, we will discuss how these challenges could be mitigated by future researchers.

For the expository task, conversely, we saw no differences across either language group on productivity. There were also no differences in modality due to the age group of the participant. However, in the monolingual group, there was a significant difference between the in-person and online testing groups, in favor of the in-person group, on the mean length of C-unit in morphemes (MLCUm), which is a measure of syntactic complexity and language proficiency. This was an interesting finding since the expository task is highly structured, and the most likely to produce complex language. However, we speculate that this difference may be due to the adaptation of the task to the online format. To the best of our knowledge, the "Favorite game or sport task" has not been administered online in a research study previously. While some aspects of this task were easy to adapt (e.g., asking the questions about their favorite sport), others were more complex. In the in-person version of this task, students are provided with optional "prompt" cards in a randomized order, which they can refer to, use as a physical manipulative, or ignore entirely. In the online version, we created several randomized versions of these cards which would appear on the screen in front of the child. However, their

positionality on the screen seemed to give the cards more weight, and almost all of the children used them throughout this task. This could, in turn, make the task more formulaic since the students were simply responding to each prompt individually. In person, children were more expansive and creative with their descriptions. Another possibility is the impact of cognitive load. Expository discourse is already a complex task that requires organization, explicit instructions, and complex language. It is possible that children struggle to complete more complex tasks online since some of their cognitive load is focused on the online testing format in addition to the assessments themselves. In summary, we suspect that the syntactic complexity of the expository measure was lower online since it is the most challenging of the tasks. The high cognitive load, combined with the challenging online interaction, may have posed particular challenges. In the clinical implications section, we will discuss how this challenge could also be mitigated by future researchers.

Across both language groups and all ages, we saw no differences on the productive narrative task when comparing the in-person and online assessment groups. We speculate that this can be explained by several factors. While this task has not been previously adapted in a research study, it was simpler to move to an online format. Children would simply see the picture on their screens instead of on the desk in front of them, and the prompts were otherwise identical. This task also usually feels less like a "test" to students, and it seems easier for them to engage since it is highly imaginative. While there were also audio and Wi-Fi issues during this task, the participants seemed to lose their train of thought less since they were continuing a narrative. We can therefore state with some degree of confidence that the results of productive narrative tasks are comparable across in-person and online testing contexts.

Clinical implications

The current COVID-19 pandemic has influenced the practices of researchers and clinicians within the healthcare field, including speech-language pathologists. Historically, speech-language pathology has depended largely on in-person interactions to assess children's language abilities. However, since the beginning of the COVID-19 pandemic, the use of online assessment and treatment models have become widely implemented, often being offered as the primary method of service. For researchers and clinicians, it is important to be aware of and account for any differences that may result from assessment modality. Particularly for clinicians, the transformations of these in-person interactions to an online medium must consider which important insights can be captured during language sample analysis. Similar to previously discussed research (Taylor et al., 2014; Manning et al., 2020), the aforementioned results further support using language sample analyses gathered online for various discourse types

(i.e., conversation, narrative, and expository language) for both monolingual and bilingual speakers aged 7–12 years. Both modalities can provide researchers and clinicians with accurate and reliable information about the child and their language abilities.

However, it is also important to discuss the modifications that may need to be made as indicated by the results of this study. As previously stated, it appears that receptive vocabulary and narrative measures are more easily adaptable to the online assessment context. To successfully use a conversational assessment, we would make two recommendations. First, it is crucial that clinicians and researchers thoroughly test any Wi-Fi and audio issues, and only proceed with the assessment if a minimum threshold is met. Furthermore, it would be advantageous to ensure the child is calm, comfortable, and engaged prior to beginning the conversational task. This could mean having a relative sit with them, informally chatting before starting the task, asking guardians for topics ahead of time, etc. Otherwise, the conversational task may not be as reflective of the students' language abilities as it would be in person. For the expository task, we speculate that the "Favorite game or sport" task may be more difficult to administer online. An expository task that does not rely on prompt cards might be more suitable for online assessment (e.g., explaining how to make a peanut butter and jelly sandwich). Alternatively, prompt cards could be eliminated for all modalities. Finally, clinicians may notice qualitative differences in online testing, including interruptions due to technical errors, lack of tactile information, or interjections from family members.

Limitations and future directions

Research comparing in-person and online assessment is relatively new in the field of speech-language pathology. The current study found that monolingual children differed on conversational productivity and expository syntactic complexity. Future studies may want to investigate why MLCUm was more sensitive to differences than other syntactic complexity measures (the number of lexical verbs by the number of C-units). Our research focused on three discourse measures for typically developing children aged 7 to 12 years of age. However, the distribution across conditions was uneven, with a higher number of students in person than those who were tested online. Future studies may want to expand on our results and focus on other populations or discourse measures to further confirm if online and in-person assessment can be used interchangeably. For example, future studies may want to include a sample of younger children, or children with language or learning difficulties, such as a Developmental Language Disorder. Our results comparing in-person and online testing may vary for other populations of children. This is especially true for the expository and conversational measures, where any

differences in modality may be exaggerated in non-typically developing populations. Furthermore, the current findings should be validated and supplemented by studies using a within-subject design. Additionally, future studies could include independent measures of the children's language and cognitive abilities. In the same vein, future studies may want to look at other microstructure or macrostructure measures. Finally, more research is needed on the impact of bilingualism on assessment modality. While no differences were found in this paper, it would be beneficial to replicate these results with sequential bilingual children and bilingual students from other language backgrounds.

Conclusion

In conclusion, it is possible to conduct measures of discourse online with similar results to that of in-person language sampling data for monolingual and bilingual children aged 7–12 years. The evidence of this study suggests that receptive vocabulary and narrative measures are reliable assessments to be used in an online context. Conversational measures may be comparable with the aforementioned safeguards in place (e.g., audio, Wi-Fi, situating the child). Expository measures should be used with caution until further research has explored the differences between modalities. While pivoting to online services can be difficult for researchers and clinicians, language sampling is a valuable resource and requires little materials for both monolingual and bilingual children. Based on the results of the current study, researchers and clinicians can feel confident in continuing to use language sampling as an informative assessment tool in the provision of online services.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Dalhousie University Research Ethics Board: Health Sciences. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

References

Berman, R., and Slobin, D. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Author contributions

DB MB, and EK-RB were responsible for the introduction. DB and EK-RB were responsible for materials and methods. DB, SR, XC, BS, and EK-RB were responsible for the results. DB, VB, KB, and JM were responsible for the discussion. All authors were involved in the revision and DB was responsible for all final edits. All authors were involved in the design of this study, in addition to the acquisition, analysis, and interpretation of data for the work. Furthermore, all authors were involved in drafting the work and revising it critically for important intellectual content. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Berman, R. A., and Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: a developmental paradox. *Discourse Process* 43, 79–120. doi: 10.1080/01638530709336894

Bliss, L. S. (2002). *Discourse Impairments: Assessment and Intervention Applications*. Boston: Allyn and Bacon.

Brydges, C. R., and Gaeta, L. (2019). An introduction to calculating Bayes factors in JASP for speech, language, and hearing research. *J. Speech Lang. Hear. Res.* 62, 4523–4533. doi: 10.1044/2019_JSLHR-H-19-0183

Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., et al. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *Am. J. Speech-Lang. Pathol.* 20, 315–330. doi: 10.1037/e584752012-090

Chenneville, T., and Schwartz-Mette, R. (2020). Ethical considerations for psychologists in the time of COVID-19. *Am. Psychol.* 75, 644. doi: 10.1037/amp0000661

Craig, H. K., Washington, J. A., and Thompson-Porter, C. (1998). Average C-unit lengths in the discourse of African American children from low-income, urban homes. *J. Speech Lang. Hear. Res.* 41, 433–444. doi: 10.1044/jslhr.4102.433

Cummins, J. (2000). *BICS and CALP. Encyclopedia of Language Teaching and Learning*, 76–79. Clevedon: Multilingual Matters.

Dewaele, J. M., and Pavlenko, A. (2003). "Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects", in *The Effects of the Second Language on the First*, eds V. Cook (Clevedon: Multilingual Matters). 120–141. doi: 10.21832/9781853596346-009

d'Orville, H. (2020). COVID-19 causes unprecedented educational disruption: is there a road towards a new normal?. *Prospects* 49, 11–15. doi: 10.1007/s11125-020-09475-0

Dunn, L., and Dunn, D. (1997). *Peabody Picture Vocabulary Test-3rd Edition*. Circle Pines, MN: American Guidance Systems. doi: 10.1037/t15145-000

Dunn, L., and Dunn, L. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., and Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Bloomington, MN: Pearson Assessments. doi: 10.1037/t15144-000

Eisenberg, A. R. (1985). Learning to describe past experiences in conversation. *Discourse Process* 8, 177–204. doi: 10.1080/01638538509544613

Eisenberg, S. L., Fersko, T. M., and Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children. *Am. J. Speech Lang. Pathol.* 10, 323–342. doi: 10.1044/1058-0360(2001/028)

Eriks-Brophy, A., Quittenbaum, J., Anderson, D., and Nelson, T. (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clin. Ling. Phonetics* 22, 589–609. doi: 10.1080/02699200802221737

Fong, R., Tsai, C. F., and Yiu, O. Y. (2021). The implementation of telepractice in speech language pathology in Hong Kong during the COVID-19 pandemic. *Telemed. E Health* 27, 30–38. doi: 10.1089/tmj.2020.0223

Gillam, R. B., and Pearson, N. A. (2017). *Test of Narrative Language–Second Edition (TNL-2)*. Austin, TX: Pro-Ed.

Guiberson, M., Rodríguez, B. L., and Zajacova, A. (2015). Accuracy of telehealth-administered measures to screen language in Spanish-speaking preschoolers. *Telemed. e-Health.* 21, 714–720.

Haaf, R., Duncan, B., Skarakis-Doyle, E., Carew, M., and Kapitan, P. (1999). Computer-based language assessment software: the effects of presentation and response format. *Lang. Speech Hear. Serv. Sch.* 30, 68–74. doi: 10.1044/0161-1461.3001.68

Heilmann, J., and Malone, T. O. (2014). The rules of the game: Properties of a database of expository language samples. *Lang. Speech Hear. Serv. Sch.* 45, 277–290. doi: 10.1044/2014_LSHSS-13-0050

Heilmann, J. J., Miller, J. F., and Nockerts, A. (2010). Using language sample databases. *Lang. Speech Hear. Serv. Schools* 41, 84–95. doi: 10.1044/0161-1461(2009/08-0075)

Hoff, E. (2009). "Language development at an early age: Learning mechanisms and outcomes from birth to five years," in *Encyclopedia on early childhood development*, 1–5.

Hughes, D., McGillivray, L. R., and Schmidek, M. (1997). *Guide to Narrative Language: Procedures for Assessment*. Eau Claire, WI: Thinking Publications.

IBM (2021). *Bayesian independent sample inference*. Available online at: https://www.ibm.com/docs/ro/spss-statistics/25.0.0?topic=statistics-bayesian-independent-sample-inference#fnsrc_1 (accessed March 06, 2022).

Imran, N., Zeshan, M., and Pervaiz, Z. (2020). Mental health considerations for children and adolescents in COVID-19 Pandemic. *Pak. J. Med. Sci.* 36, S67. doi: 10.12669/pjms.36.COVID19-S4.2759

Jia, G., Chen, J., Kim, H., Chan, P. S., and Jeung, C. (2014). Bilingual lexical skills of school-age children with Chinese and Korean heritage languages in the United States. *Int. J. Behav. Dev.* 38, 350–358. doi: 10.1177/0165025414533224

Kay-Raining Bird, E., Joshi, N., and Cleave, P. L. (2016). Assessing the reliability and use of the expository scoring scheme as a measure of developmental change in monolingual English and bilingual French/English children. *Lang. Speech Hear. Serv. Sch.* 47, 297–312. doi: 10.1044/2016_LSHSS-15-0029

Lavigne-Cerván, R., Costa-López, B., Juárez-Ruiz de Mier, R., Real-Fernández, M., Sánchez-Muñoz de León, M., and Navarro-Soria, I. (2021). Consequences of COVID-19 confinement on anxiety, sleep and executive functions of children and adolescents in Spain. *Front. Psychol.* 12, 334. doi: 10.3389/fpsyg.2021.565516

Le Normand, M. T., Parisse, C., and Cohen, H. (2008). Lexical diversity and productivity in French preschoolers: developmental, gender and sociocultural factors. *Clin. Ling. Phonet.* 22, 47–58. doi: 10.1080/02699200701669945

Leadholm, B., and Miller, J. F. (1992). *Language Sample Analysis: The Wisconsin Guide*. Madison, WI: Department of Public Instruction.

Manning, B. L., Harpole, A., Harriott, E. M., Postolowicz, K., and Norton, E. S. (2020). Taking language samples home: Feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *J. Speech Lang. Hear. Res.* 63, 3982–3990.

Mansuri, B., Tohidast, S. A., Mokhlesin, M., Choubineh, M., Zarei, M., Bagheri, R., et al. (2021). Telepractice among speech and language pathologists: a KAP study during COVID-19 pandemic. *Speech Lang. Hear.* 1–8. doi: 10.1080/2050571X.2021.1976550. [Epub ahead of print].

Manzanares, B., and Kan, P. F. (2014). assessing children's language skills at a distance: does it work? *Perspect. Augment. Altern. Commun.* 23, 34–41. doi: 10.1044/aac23.1.34

Merriam-Webster. (n.d.). *Discourse. In Merriam-Webster.com dictionary*. Available online at: https://www.merriam-webster.com/dictionary/discourse (accessed October 19, 2021).

Miller, J., and Chapman, R. (2012). *Systematic Analysis of Language Transcripts (SALT)[Computer software]*. Middleton, WI: SALT Software LLC.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., and Francis, D. J. (2006). Oral language and reading in bilingual children. *Learn. Disabil. Res. Pract.* 21, 30–43. doi: 10.1111/j.1540-5826.2006.00205.x

Nippold, M. A. (2009). School-age children talk about chess: does knowledge drive syntactic complexity? *J. Speech Lang. Hear. Res.* 52, 856–871. doi: 10.1044/1092-4388(2009/08-0094)

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., and MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: examining the use of complex syntax. *J. Speech Lang. Hear. Res.* 57, 876–886. doi: 10.1044/1092-4388(2013/13-0097)

Nippold, M. A., Hesketh, L. J., Duthie, J. K., and Mansfield, T. C. (2005). Conversational versus expository discourse: a study of syntactic development in children, adolescents, and adults. *J. Speech Lang. Hear. Res.* 48, 1048–1064. doi: 10.1044/1092-4388(2005/073)

Nippold, M. A., Mansfield, T. C., and Billow, J. L. (2007). Peer conflict explanations in children, adolescents, and adults: examining the development of complex syntax. *Am. J. Speech Lang. Pathol.* 16, 179–188. doi: 10.1044/1058-0360(2007/022)

Nippold, M. A., and Sun, L. (2010). Expository writing in children and adolescents: a classroom assessment tool. *Perspect. Lang. Learn. Educ.* 17, 100–107. doi: 10.1044/lle17.3.100

Orgilés, M., Espada, J. P., Delvecchio, E., Francisco, R., Mazzeschi, C., Pedro, M., et al. (2021). Anxiety and depressive symptoms in children and adolescents during covid-19 pandemic: a transcultural approach. *Psicothema* 33, 125–130. doi: 10.7334/psicothema2020.287

Owens, R. (2012). *Language Development: An Introduction, 8th Edn*. Boston, MA: AUyn and Bacon.

Paul, R., and Norbury, C. F. (2012). Language disorders from infancy through adolescence. St Louis, MI: Elsevier Health Sciences.

Putri, R. S., Purwanto, A., Pramono, R., Asbari, M., Wijayanti, L. M., and Hyun, C. C. (2020). Impact of the COVID-19 pandemic on online home learning: an explorative study of primary schools in Indonesia. *Int. J. Adv. Sci. Technol.* 29, 4809–4818. Available online at: http://sersc.org/journals/index.php/IJAST/article/view/13867

Reimers, F., Schleicher, A., Saavedra, J., and Tuominen, S. (2020). Supporting the continuation of teaching and learning during the COVID-19 pandemic. *OECD* 1, 1–38. Available online at: https://globaled.gse.harvard.edu/files/geii/files/supporting_the_continuation_of_teaching.pdf

Burchell et al. 10.3389/fcomm.2022.798196

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., and Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *J. Speech Lang. Hear. Res.* 52, 333–349. doi: 10.1044/1092-4388(2009/08-0183)

Schneider, P., Dubé, R. V., and Hayward, D. (2004). *The Edmonton Narrative Norms Instrument*. Available online at: http://www.rehabmed.ualberta.ca/spa/enni/ (accessed October 19, 2021).

Stadler, M. A., and Ward, G. C. (2005). Supporting the narrative development of young children. *Early Childh. Educ. J.* 33, 73–80. doi: 10.1007/s10643-005-0024-4

Taylor, O. D., Armfield, N. R., Dodrill, P., and Smith, A. C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *J. Telemed. Telecare* 20, 405–412. doi: 10.1177/1357633X14552388

Turkstra, L. S., Quinn-Padron, M., Johnson, J. E., Workinger, M. S., and Antoniotti, N. (2012). In-person versus telehealth assessment of discourse ability in adults with traumatic brain injury. *J. Head Trauma Rehabil.* 27, 424.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., et al. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon. Bull. Rev.* 28, 813–826. doi: 10.3758/s13423-020-01798-5

Verhoeven, L., and Strömqvist, S. (Eds.). (2001). *Narrative Development in a Multilingual Context*, Vol. 23. Amsterdam: John Benjamins Publishing. doi: 10.1075/sibil.23

Westerveld, M. F., and Moran, C. A. (2013). Spoken expository discourse of children and adolescents: retelling versus generation. *Clin. Ling. Phonet.* 27, 720–734. doi: 10.3109/02699206.2013.802016

Zebib, R., Tuller, L., Hamann, C., Abed Ibrahim, L., and Prévost, P. (2020). Syntactic complexity and verbal working memory in bilingual children with and without developmental language disorder. *First Lang.* 40, 461–484. doi: 10.1177/0142723719888372

Appendix

TABLE A1 Summary of all results by population, measure and analysis.

Population, measure, and analysis	Dependent variables	Results
Monolinguals		
Vocabulary, ANOVA	PPVT Raw score	Modality, NS
		Age[*]
		Modality × Age, NS
Conversation, MANOVA	MLCUm	Modality, NS
		Age, NS
		Modality × Age, NS
	NTW	Modality[*]
		Age[*]
		Modality × Age, NS
	SC	Modality, NS
		Age, NS
		Modality × Age, NS
Expository, MANOVA	MLCUm	Modality[*]
		Age[*]
		Modality × Age, NS
	NTW	Modality, NS
		Age[*]
		Modality × Age, NS
	SC	Modality, NS
		Age, NS
		Modality × Age, NS
Narration, MANOVA	MLCUm	Modality, NS
		Age, NS
		Modality × Age, NS
	NTW	Modality, NS
		Age, NS
		Modality × Age, NS
	SC	Modality, NS
		Age[*]
		Modality × Age, NS
Simultaneous bilinguals		
Vocabulary, ANOVA	PPVT Raw score	Modality, NS
		Age[*]
		Modality × Age, NS

(Continued)

TABLE A1 Continued

Population, measure, and analysis	Dependent variables	Results
Conversation, MANOVA	MLCUm	Modality, NS
		Age[*]
		Modality × Age, NS
	NTW	Modality, NS
		Age[*]
		Modality × Age, NS
	SC	Modality, NS
		Age[*]
		Modality × Age, NS
Expository, MANOVA	MLCUm	Modality, NS
		Age[*]
		Modality × Age, NS
	NTW	Modality, NS
		Age[*]
		Modality × Age, NS
	SC	Modality, NS
		Age, NS
		Modality × Age, NS
Narrative, MANOVA	MLCUm	Modality, NS
		Age[*]
		Modality × Age, NS
	NTW	Modality, NS
		Age[*]
		Modality × Age, NS
	SC	Modality, NS
		Age[*]
		Modality × Age, NS

PPVT, Peabody Picture Vocabulary Test, 4th edition (Dunn and Dunn, 2007); MLCUm, mean length of communication unit in morphemes; NTW, number of total words; SC, syntax complexity; NS, not significant.

[*]Denotes statistical significance.