Top Lang Disorders
Vol. 43, No. 4, pp. 349–366
Copyright © 2023 Wolters Kluwer Health, Inc. All rights reserved.

A Feasibility Study in Virtual Assessment Procedures of a Sentence-Writing Probe for Use With Intermediate-Grade Students

Karissa J. Marble-Flint and Anthony D. Koutsoftas

This article reports on the development and initial feasibility of virtual assessment procedures for a sentence-writing probe for remote instructional purposes with intermediate-grade students. The study included a sample of 15 intermediate-grade children. The sentence-writing probe was administered through video conferencing software, an innovation of the times, across three sessions separated by 2 weeks. Scores derived from sentence probes included total number of words, a sentence accuracy score, and a word accuracy score, which were compared across time points. Results indicated no statistically significant differences across time points for the entire sample for all measures except the total number of words at Time 2. Measures obtained from the sentence-writing probe were significantly correlated with standardized measures of oral language. Findings from this study support the proof of concept that virtual assessment procedures can be used to assess sentence-level writing in intermediate-grade students. Future directions are provided regarding the utility of remote instruction for assessment purposes, the types of scores derived from measures, and future plans to scale up the assessment for use in research studies and as a curriculum-based evaluation tool. **Key words:** *assessment*, *intermediate-grade students*, *levels of language*, *remote instruction*, *writing*

Author Affiliations: *Department of Communication Sciences & Disorders, Wichita State University, Wichita, Kansas (Dr Marble-Flint); and Department of Speech Language Pathology, Interprofessional Health Sciences (IHS) Campus, Seton Hall University, Nutley, New Jersey (Dr Koutsoftas).*

This work was supported by an award from the American Speech-Language-Hearing Association (ASHA) Advancing Academic-Research Careers (AARC) awarded to Marble-Flint. This work was partially supported through a development and innovation grant from the Institute of Education Sciences, Grant #R324A200046, awarded to Koutsoftas. The opinions expressed are those of the authors and do not represent views of either ASHA or the U.S. Department of Education. In addition, ASHA and the U.S. Department of Education did not have any involvement in study design; in the collection, analysis and interpretation of data; in the writing of the reports; or in the decision to submit the article for publication.

The author and planners have disclosed no potential relevant financial relationships or otherwise. Author disclosures can be found at http://links.lww.com/TLD/A112.

INTRODUCTION

Documenting changes in schoolchildren's writing is an important consideration for both researchers and clinicians. One common way to document changes in writing is to obtain writing samples and analyze them using measures at multiple levels of language

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.topicsinlanguagedisorders.com).

Corresponding Author: *Karissa J. Marble-Flint, PhD, CCC-SLP, Department of Communication Sciences & Disorders, Wichita State University, 1845 Fairmount St, Wichita, KS 67260-0075 (karissa.marble-flint@wichita.edu).*

DOI: 10.1097/TLD.0000000000000322

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

(word, sentence, discourse) following acceptable language transcription methods (e.g., Dockrell et al., 2014; Koutsoftas, 2018; Puranik et al., 2008). Such an approach does well to provide a representative portrayal of writing ability; however, not all writing interventions target discourse-level writing, so the use of a prompt to elicit discourse-level samples may not be well matched to interventions that target word- or sentence-level skills. Innovations in technology have contributed to an increase in the social acceptance and use of virtual or remote instructional practices, including the use of video conferencing software by schoolchildren and caregivers for academic tasks. Yet, there is a dearth of research supporting virtual assessment procedures, especially for writing skills, in schoolchildren. This article reports on the development and feasibility of a sentence-writing probe administered remotely that can be used as part of writing interventions for intermediate-grade students (i.e., fourth and fifth graders).

Theoretical framework for writing

Writing is a complex cognitive–linguistic process that requires the orchestration of multiple skills to represent language in written form (García & Fidalgo, 2008). The simple view of writing (Berninger & Amtmann, 2003) is a widely accepted theoretical framework used to describe writing and depicts the interrelationships of text generation, transcription, and self-regulation skills needed for writing texts, all centered around working memory. Text generation is the generation of words, sentences, and paragraphs, and occurs subvocally or aloud. Text generation is *oral language* that has yet to be transcribed. Transcription skills include spelling, handwriting, or keyboarding skills needed to transcribe oral language into written language. Self-regulation skills include executive functions such as planning, revising, goal setting, and theory of mind (or perspective taking) needed to engage in writing purposeful and meaningful texts.

In the current study, we considered the simple view of writing (Berninger & Amtmann, 2003) as the launching point for developing a virtual assessment of sentence-level writing skills for intermediate-grade schoolchildren. There were two additional rationales for the development of this virtual writing assessment. First was the need to document changes in sentence-level writing as part of intervention research, in particular, single-case research designs. Second was the need to obtain a representative writing sample while providing students with scaffolds that support the level of language being assessed (e.g., word, sentence) and simultaneously considering developmental expectations for students.

Documenting progress for research purposes

The primary motivation for developing this sentence-writing virtual assessment was the need to document progress as part of single-case research designs. Single-case designs are well suited to demonstrate the feasibility and early efficacy of an intervention (Rogers & Graham, 2008) and for interventions designed for low incidence or highly heterogeneous (e.g., autism spectrum disorder, hearing impaired) populations (e.g., Fey & Finestack, 2009; Kratochwill & Levin, 2014; Olswang, 1998). Best practices for single-case designs require that the target behavior is measured with regularity over the course of the intervention (Kratochwill et al., 2013), oftentimes referred to as probes. For example, over the course of a 6-week writing intervention using a single-case research design, one might attempt to elicit a discourse-level writing sample each week of the study; however, this is a problem for two reasons. First, it may be the case that the writing intervention targets word- or sentence-level writing skills and not discourse-level writing, so the use of discourse-level written elicitation procedures would not be well matched to the intervention. Second, obtaining a discourse-level writing sample in response to a prompt can be burdensome for research participants,

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

especially those who come from special populations and are struggling with writing. In the current study, we designed sentence-level writing probes that ask students to generate five sentences each time, which aligns with recommendations for single-case research designs (Kratochwill et al., 2013). Probes can be included weekly either at the beginning of the first session or at the end of the last session per week of intervention.

Scaffolded writing assessment

The secondary motivation for developing this sentence-writing probe task was to create a task that assists students with generating a more representative writing sample. Although scaffolds are common practice in language interventions, it was important that the scaffolds included in the current study aligned with current and widely accepted theoretical frameworks for writing (e.g., Berninger & Amtmann, 2003). Ritchey et al. (2016) suggested two additions to the simple view of writing to support developing writers. First was the inclusion of a levels of language framework whereby one must consider written language output at the subword, word, sentence, and discourse levels. This is something that has been well supported by research (e.g., Abbott et al., 2010; Berninger et al., 2010). Some students are writing at the subword or word levels whereas others may be writing at the sentence or discourse levels. Thus, interventions may be designed to target one or more skills within or across levels. Second was the inclusion of scaffolds that are visual, verbal, and written to support students when writing. Because of the demands of working memory needed for writing (Berninger & Amtmann, 2003; McCutchen, 1996; Vanderberg & Swanson, 2007), young writers and those who struggle benefit from scaffolds to ensure optimal writing output. Scaffolds can account for developmental or individual variation in writing skills. Developmental scaffolds account for age- or grade-level expectations when developing writing tasks, whereas individual scaffolds account for intra- and interindividual differences. For example, some students can generate ideas for writing orally as part of text generation; however, due to constraints of working memory, the final written sample may be degraded. Thus, having students rehearse a segment of text before writing can help them generate more complete and accurate writing. Scaffolds can provide support not only for the development of skills as part of interventions but also when obtaining or eliciting writing samples for assessment or progress-monitoring purposes.

To obtain a more representative writing sample, one must account for developmental and individual variations as much as possible and scaffolds can achieve this purpose (Graham, 1990; McCutchen, 1996). For example, visual scaffolds can include pictures to support the generation of ideas and verbal scaffolds can include generating ideas aloud before transcribing text, referred to as a *verbal rehearsal strategy*. Written scaffolds may include key words or phrases provided in writing to support word generation and spelling as part of the transcription process. For example, providing written key words a student might be expected to include in their writing helps ensure that text generation is not constrained by transcription. Likewise, providing a familiar topic helps ensure that background knowledge does not constrain text generation (and visual scaffolds using pictures can further support this). As an example of a written scaffold, if a child is asked to write a sentence that includes a specific word such as a cohesive tie (e.g., besides, thus, however, similarly), the cohesive tie is provided to them in writing as a written scaffold. If the child writes a sentence that includes the cohesive tie but does not use the word correctly, then text generation and transcription are accounted for, and one can conclude that the child does not know the meaning or grammatical use of the word.

Prior research supports the use of a variety of scaffolds to support developing writers when eliciting writing samples. McMaster et al. (2009) reported on the technical adequacy of a series of curriculum-based

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

measurement probes administered to a sample of first and second graders. Story and photo prompt tasks emerged as most promising with respect to reliability, validity, and stability over time. For the story prompt, a sentence starter was provided in writing and read aloud to students; for the photo prompt, a relatable photo was provided for which students generated a story. Across both tasks, students were provided time to plan their stories before writing. Arfé and Pizzocaro (2016) compared an oral and written sentence generation task as a way to assess written expression in Italian children in Grades 2 through 5. For the oral sentence generation task, students were given two words and asked to orally generate as many sentences as they could using the target words within 2 min. For the written sentence generation task, students were given two pairs of words in writing on a lined sheet of paper and asked to compose as many different sentences as they could within 5 min. Their findings indicated that writing performance measures obtained from both tasks improved with grade and showed an advantage for the oral sentence generation task. Ritchey and Coker (2013) investigated the utility of two curriculum-based writing probe tasks for second- and third-grade students' narrative writing. One task provided a single picture prompt whereas the other provided multiple pictures, and the latter was considered more scaffolded than the former. The authors expected that additional visual scaffolds of multiple pictures would yield significantly better writing scores; yet, this was not the case. On the one hand, this suggests that providing visual scaffolds to elicit writing samples may be supportive, but on the other hand, the increased number of pictures did not result in better writing outcomes.

Providing visual scaffolds alone may not be enough to support elicitation of a representative writing sample. Heilmann and Malone (2014) obtained spoken expository samples from students by providing planning time to think, organize, and make notes for what they wanted to speak about in their expository retell. The authors attributed the use of this type of scaffold (i.e., planning time for spoken discourse) to the success of building a database of 257 spoken expository samples. Two studies examined the effects of providing structured versus unstructured planning time when eliciting writing samples (Berninger et al., 1996; Whitaker et al., 1994). Students in structured planning time groups used a strategy or responded to questions to support the generation of ideas, whereas the unstructured planning groups simply had time to plan. Although there were no between-group differences found in fourth- through sixth-grade students (Whitaker et al., 1994), there were between-group differences found in junior high school students (Berninger et al., 1996) in favor of the unstructured planning condition. That is, junior high school students who were in the unstructured planning condition received higher scores than students in the structured planning condition. Combined, these findings suggest that visual, verbal, and written scaffolds can support students in producing representative writing samples.

THE CURRENT STUDY

With increased social acceptance of virtual assessment procedures, it is important to programmatically develop measures to support an evidence base for the utility of virtual assessment procedures, especially for writing. Dually important is the need to support research participants in generating a representative sample of their writing abilities, especially schoolchildren who may be struggling as they develop writing skills. Thus, the purpose of the current study was to describe the development and feasibility of a sentence probe task designed to be administered through remote procedures (i.e., telepractice). This is the first step in development of the sentence-writing probes, with the overarching goal to have a tool that can document changes in word- and sentence-level writing over time, especially as part of single-case research designs. The probe task was developed for use through remote administration

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

using videoconferencing software (e.g., MS Teams, Zoom), an innovation we expect can support researchers and practitioners with telepractice and technological integration for future research and clinical purposes. In line with recommendations for feasibility testing (Platt & D'Anna, 2022) and scale development (Boateng et al., 2018), our research questions focused on the usability of the task by end users, intermediate-grade students, and their caregivers, as well as item development to support the content validity of the sentence-writing probes. The specific research questions were twofold. First, can the sentence-writing probe task be administered using virtual assessment procedures with regard to (a) clarity of instructions/protocols for tasks for both students and caregivers; (b) time commitment for the task; (c) the ability to recruit and retain participants for a remote assessment; (d) adherence and compliance with remote tasks; and (e) time needed to collect and analyze data? Second, can the measures obtained from the sentence-writing probes adequately describe word- and sentence-level writing proficiency in the writing samples based on stability over time and relation to one another and standardized measures of oral and written language? Third, what feedback do end users have about the virtual writing probe task?

Fourth- and fifth-grade students were purposefully selected for this study as they were expected to be able to write sentences and participate in remote assessment with minimal support from caregivers but yet show variability in their writing skills and strong adherence to the assessment task. We expected no differences in probe task performance between time points because there was no intervention provided. If we were able to demonstrate parity between probes and the words used as stimuli for each probe, then this would demonstrate feasibility of the probe assessment task for use in a research or treatment paradigm, possibly for assessing response to intervention. We predicted that data obtained from the sentence-writing probes would be significantly related to standardized sentence-level measures of oral and written language as an initial demonstration of their appropriateness for evaluating writing performance in the context of intervention.

METHOD

Study procedures were approved by the Institutional Review Board at Wichita State University (IRB#4790) and for data analysis at Seton Hall University (IRB#2021-196). Participants included fourth- and fifth-grade students who were recruited through electronic flyers, email announcements, and posts to social media. Written informed consent was obtained from parents/caregivers of the participants and the participants provided verbal assent. There were 19 participants recruited for the study; however, only 15 completed the study. Reasons for attrition included decreased interest or tolerance for online tasks ($n = 1$), missing or incorrect data transferred by parent ($n = 2$), and loss of computer privileges during the study ($n = 1$).

Participants

This study included 15 participants who met the following inclusion criteria: (1) had not been retained a grade in school, (2) had not been home-schooled except for remote schooling because of the COVID-19 pandemic, (3) primary language was English, and (4) basic writing proficiency in English. These criteria were verified using data collected from a parent intake questionnaire used in prior research and designed to obtain demographic and descriptive information about student participants. Based on parent report using the questionnaire, six participants were described as struggling writers and the remaining nine students were described as typical writers with no current or past difficulties with writing reported. This allowed for variability in the sample to allow for demonstration of feasibility of the task for a range of writers. The mean age in years for the sample was 10.47 ($SD = 0.58$), four of whom were girls. The mean years of mother's

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

education for the sample was 16.13 (SD = 1.46), and this served as our indicator of socioeconomic status.

Procedures

The research protocol included four remote sessions that were each separated by two weeks. For six participants, all four sessions overlapped with the school year, for five participants all four sessions occurred during summer break, and for the remaining four participants sessions overlapped by 1–2 weeks either with the end or the beginning of the academic year. This allowed for variability in time of administration to support the feasibility of task administration regardless of school attendance. The first and second sessions included administration of standardized assessments. These standardized assessments were administered and scored following procedures described in each test's manual with adaptations for telepractice. The second through fourth sessions included administration of the sentence-writing probes designed for the study. The research protocol was administered by graduate students in speech–language pathology who were trained and supervised by the authors of this article. Visual stimuli for standardized and experimental tasks were presented via slide decks using share screen features available in videoconferencing software. Written responses were photographed or scanned by the parent/caregiver and sent to the research team for analysis. Throughout each data collection session, the graduate students provided participants with breaks as needed. In addition, the graduate students sustained participant motivation by displaying electronic "sticker books" that were PowerPoint slides with GIFs of various animals or characters. Participants selected whether they would like to see a GIF of an animal or a character. As an incentive, participants, including those who did not complete all four study sessions, received a paper book via mail.

Standardized measures

Participants completed the Formulated Sentences and Recalling Sentences subtests from the Clinical Evaluation of Language Fundamentals, Fifth Edition (CELF-5; Wiig et al., 2013), one during the first session and the other during the second session. The Formulated Sentences subtest is an expressive language task that requires the participant to verbally formulate a sentence based on picture stimuli using a target word. Target words varied and included nouns, verbs, and coordinating or subordinating conjunctions. The verbal response is spoken aloud by the participant and recorded in writing by the examiner; the participant is not required to write. The Recalling Sentences subtest is a receptive language task that is also a measure of verbal working memory and requires the participant to repeat an orally presented sentence verbatim. For the CELF-5 Formulated Sentences subtest, psychometrics from the test's manual indicate good internal consistency reliability for the age range of our participants as follows: 9:0-9:11 = 0.90, 10:0-10:11 = 0.86, and 11:0-11:11 = 0.83. For the CELF-5 Recalling Sentences subtest, psychometrics indicate strong internal consistency reliability for the age range of our participants as follows: 9:0-9:11 = 0.95, 10:0-10:11 = 0.94, and 11:0-11:11 = 0.92. Strong validity for the CELF-5 is demonstrated by intercorrelations between subtests within the test and correlations of scores with a prior version of the test ranging from 0.78 to 0.92.

Participants completed the Sentence Combining subtest from the Test of Written Language, Fourth Edition (TOWL-4; Hammill & Larsen, 2009), which includes 23 items that use two to four short sentences presented in writing, from which the participant has to create one complex or compound sentence. Administration of this subtest was split across two test sessions, with sentences 1–12 administered during the first session and sentences 13–23 during the second session. For the TOWL-4 subtest, the internal consistency reliability as measured through Cronbach's coefficient α scores ranges from 0.74 to 0.92 across ages (McCrimmon & Climie, 2011). The test's manual indicates moderate correlations of the subtest with related measures of literacy as a demonstration of validity.

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Sentence-writing probe task

The sentence-writing probe task was administered during the second through fourth remote sessions. Each sentence-writing probe session started with a demonstration by the researcher and one or two practice opportunities for the participant before the child completed five sentence probes used for analysis. The graduate students presented stimuli via PowerPoint whereby each slide included a picture, one single-clause sentence written on the slide, and the target word the participant had to use to create their own sentence. The sentence-writing probe was modeled after procedures from the Formulated Sentences subtest of the CELF-5 (Wiig et al., 2013), which is a spoken language task that does not require the participant to write. We adapted the task to include a written component to align with writing assessment and instruction as well as to provide scaffolds to support planning and speaking sentences aloud before writing.

The sentence-writing probe task developed for this study provided participants with scaffolds to support writing output, aligned with the Ritchey et al. (2016) model of writing assessment that includes visual, verbal, and written scaffolds. Visual scaffolds were provided to participants through the use of vivid and relatable pictures of activities such as walking in the rain, playing at the playground, or helping in the kitchen. Pictures used in the task were selected to reflect relatable events that included images of children and were available from copyright-free photo websites (e.g., Pixabay, Unsplash, Flickr). Written scaffolds included a single-clause sentence about the picture along with the target word provided in writing that the participant was expected to use. Single-clause sentences about the picture were in subject/verb/object format and included key words related to the picture, and each sentence started with the word "the." This ensured that text generation would not be limited by transcription or working memory challenges. Verbal scaffolds were provided when the student was asked to verbalize the sentence before writing, referred to here and throughout as a *verbal rehearsal strategy*. The verbal rehearsal strategy used to elicit writing samples was adapted from prior research (Berninger, 2009) and was as follows: think it, say it, and write it. During the demonstration, the researcher modeled the task using the following script, with nonverbal instructions provided in brackets. An example of stimuli used for this demonstration is available as a Supplemental Digital Content appendix, available at: http://links.lww.com/TLD/A108.

"Here is a picture of a little girl crying. The sentence says, 'The girl feels sad.' I will make a new sentence about this picture using the word *because*. [Wait five seconds before speaking to demonstrate thinking time.] My new sentence is, 'The girl is crying because she is sad.' Now I will write it down. [Demonstrate writing down the sentence.] I wrote, 'The girl is crying because she is sad.'"

Following the researcher's demonstration, participants completed practice trials to ensure that they understood the task. Participants completed two trials during the first sentence-writing probe session and one trial for each remaining two sentence-writing probe sessions. First, the researcher asked the participant to think of a sentence using the target word. Next, the child verbally rehearsed their sentence using the selected target word. At the same time, the researcher wrote the participant's verbal rehearsal verbatim onto a response form. Then, the participant wrote the sentence on a piece of paper provided by their parent/caregiver. Finally, the participant read the sentence they wrote. Again, the researcher recorded the response onto a response form. At the conclusion of the research session, the caregiver emailed the research team a scanned copy or photograph of the participant's written sentences. An example of the script for the actual sentence-writing probe is provided below. The first target word was a familiar noun, *dinosaur*, and included a picture of a boy playing with toy dinosaurs.

"This sentence says, 'The boy plays with his new toys.' Now you make a sentence about

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

this picture using the word *dinosaur*. Make sure the sentence is about the picture and uses the word *dinosaur*. When you are ready, tell me your sentence. [Allow child 10 seconds to respond. If they do not speak on their own, then prompt further by repeating the instructions. Write down what they say. After the student verbalizes the sentence, continue with this script.] Now write it down. [Allow time to write.] Read me what you wrote. [Write down what the student read back to you.]"

Target word selection

There were 15 target words used in the sentence-writing probes, five for each day, plus an additional seven words used for demonstrations and practice trials. Cohesive ties were selected as these are important words for intermediate-grade students to use when writing sentences reflecting complex ideas (Koutsoftas & Petersen, 2017). Koutsoftas and Petersen were able to estimate the frequency of cohesive ties in writing samples from intermediate-grade students, and the most frequently reported cohesive ties from that study were included in the current study, randomized across time points. Cohesive ties include coordinating and subordinating conjunctions across the categories of additive, causal, coordinating, temporal, and adversative. The task purposefully began with nouns and verbs to familiarize the participants with the procedure and as a way to gauge basic writing skills. If students were not yet able to write sentences using cohesive ties, then at minimum the task allowed for observation of simple sentence generation using nouns or verbs. See the Appendix for target words in order of administration for each probe.

Sentence-writing probe measures

There were three scores obtained for each of the sentences composed by participants from the sentence-writing probes. These were (a) the total number of words (TNW), (b) a sentence accuracy score (SAS), and (c) a word accuracy score (WAS); the latter two were developed for the study. Sentence scoring was completed by undergraduates in education or speech–language pathology who completed a 2-hr training for this task. They typed verbatim all 15 sentences into one document for each participant, retaining capitalization, spelling, grammar, and punctuation as produced by the student. Any challenges with legibility were addressed by consulting the printed record form or the authors. The verbatim typed transcripts were used for subsequent analyses.

Total number of words

The TNW was calculated by counting the number of words produced in the sentence. This measure by itself was explored in later analyses; however, it was also used to control for the length of sentences for the SAS.

Sentence accuracy score

An SAS was calculated by first analyzing sentences for correct word sequences (CWS) and incorrect word sequences (IWS). CWS accounted for subword-, word-, and sentence-level written language output and has been shown to be sensitive to developmental and instructional changes (McMaster et al., 2009, 2011). CWS scoring accounts for spelling, grammar, capitalization, and punctuation for each set of two adjacent words within a sentence. Likewise, IWS scoring accounts for two adjacent words that are incorrect in terms of spelling, grammar, capitalization, or punctuation. To illustrate, if two adjacent words have no spelling, grammar, capitalization, or punctuation errors, then it is considered a CWS and is awarded a CWS score of 1 and the IWS is scored as zero. Likewise, if the sequence is considered incorrect, then the CWS is scored zero and the IWS is scored as a 1. This scoring starts at the beginning of the sentence whereby only the first word is evaluated for spelling, grammar, and capitalization. From there, the first and second words are evaluated, and this continues for every two words until the end of the sentence is reached. The space between the last word and final punctuation is considered as a word sequence so that final punctuation

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

can be evaluated in the metric. CWS is generally applied to samples obtained in a set amount of time or under similar conditions so that the length of writing sample is accounted for within and across participants. For the current study, we controlled for length by dividing the IWS by the TNW for each sentence. We selected the IWS because it was unlikely that the IWS would be of greater value than the TNW. In contrast, the CWS could be one point higher than the TNW for sentences that were produced without errors. This measure is referred to as the SAS and was calculated using the following formula: [1 − (IWS/TNW)]. We subtracted from 1 so that higher numbers represented greater sentence accuracy.

Word accuracy score

The third score applied to each sentence was a WAS. The scoring system was a 3-point scale that accounts solely for accurate use of the target word regardless of sentence accuracy. A score of 0 indicated that the target word was used incorrectly. A score of 1 indicated that the target word was used correctly but not in the manner expected. For example, if the target word "so" was used to start a sentence but did not indicate a causal relationship between two clauses, then a score of 1 was applied. A score of 2 indicated that the target word was used correctly and as intended.

Reliability

Interrater reliability was calculated on 20% of standardized tests and experimental measures whereby a second graduate or undergraduate trained on the scoring or coding procedures independently scored for comparison with the primary scorer. For standardized tests, interrater agreement was calculated for the two subtests of the CELF-5 and the Sentence Combining subtest of the TOWL-4 using point-to-point agreement for each item. For the CELF-5, interrater agreement for both Recalling Sentences and Formulated Sentences was 100%, and for the Sentence Combining subtest of the TOWL-4 interrater agreement was 91.74%. For experimental measures, point-to-point agreement was calculated for three participants' full data sets, which included 15 sentences included in analyses, with interrater agreement as follows: TNW = 99.78%, CWS = 97.99%, IWS = 88.28%, and WAS = 88.89%.

Follow-up interview

To obtain end user feedback from caregivers, a follow-up interview was completed individually using the virtual platform. The purpose of the interview was to gather suggestions for enhancing the writing probe assessment task, which was explained with caregivers at the onset of the interview. Given that this was a feasibility study, we were primarily concerned with the caretakers' level of participation as a variable of interest.

Each interview lasted 15 min or less. Completion of the interview was not required; all 15 caregivers participated. The interview included five questions about caregivers' level of agreement using a five-point rating scale. Interview questions were displayed on the screen using the share screen feature, which allowed caregivers to reference the scale as they were answering the questions. Table 1 includes follow-up interview questions and percentages of responses. The sixth question was open-ended and asked caregivers to provide feedback in response to their experience with the study.

RESULTS

Research question 1

The first research question asked whether the sentence-writing probe could be administered using virtual assessment procedures with intermediate-grade students and their caregivers with regard to five characteristics. First, the clarity of instructions was demonstrated by 15 participants being able to complete the entire study, including transfer of written responses by caregivers via text or email. Second, the time commitment for the probe assessment task for student participants was approximately 15 min per session,

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Table 1. Frequency of response to Likert rating scale questions from parents/caregivers ($N = 15$)

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Instructions and procedures for the study were clear and easy to follow, including accessing videoconferencing software, initial online survey, and research study sessions.	93%		7%		
2. Communications in writing and speaking from the graduate student clinician were clear and easy to understand.	93%		7%		
3. The frequency of communication with the graduate student clinician and the researcher was appropriate for the purposes of the research study.	93%		7%		
4. Overall, the remote writing study was clear and easy for me and my child to participate in.	93%		7%		
5. My child enjoyed participating in the remote writing study.	27%	47%	27%		

spaced two weeks apart. The time commitment for caregivers included a 10-min orientation at the beginning of the study and differed depending on the level of monitoring needed by participants. All participants were able to use the remote technology independently within the first session. Third, regarding attrition, while we were able to recruit 19 participants for the study, four of them did not complete the study for various reasons. One expected reason was a lack of tolerance for the online nature of the study as was the case with one participant, one unexpected reason was the loss of computer privileges imposed by parent or caregiver that did not allow one participant to continue. The remaining reason for attrition was the loss of data transfer by parents or caregivers, which is an important outcome of this feasibility study. Fourth, participants were able to adhere to task demands as indicated by completion of tasks by participants who completed the study. This was supported by the establishment of rapport between each participant and the examiner, breaks as needed, and the use of a virtual sticker book described earlier. Fifth, the time needed to collect all data included four 45-min sessions for each participant conducted individually by trained assistants plus the time needed to complete the scoring procedures, which we estimated at 60 min per sample and included scoring standardized and experimental measures across all four time points. Combined, this indicates approximately 5 hr per participant enrolled in the study, distributed across time points.

Research question 2

The second research question asked whether the measures obtained from the sentence-writing probe could be used to describe word- and sentence-level writing in the sample. To start, we provide descriptive data for standardized and experimental measures, averaged across the sample of 15 participants. For the CELF-5 subtests, the mean subtest standard score for Formulated Sentences was 11.60 ($SD = 3.18$) and for Recalling Sentences was 10.33 ($SD = 3.46$). For the Sentence Combining subtest of the

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Table 2. Means and standard deviations for experimental measures across the total sample ($N = 15$) for each time point and across total sentences produced

Measure	Time 1	Time 2	Time 3	Total
Total number of words	9.33 (1.83)	10.43 (0.84)*	9.56 (1.12)	9.70 (1.00)
Sentence accuracy score	0.74 (0.22)	0.78 (0.15)	0.76 (0.19)	0.76 (0.17)
Word accuracy score	1.45 (0.28)	1.41 (0.28)	1.57 (0.49)	1.48 (0.24)

*$p < .05$.

TOWL-4, the mean standard score for the sample was 12.93 ($SD = 4.54$). Table 2 includes means and standard deviations for experimental measures including the TNW, SAS, and WAS, across the total sample for each time point and across all 15 sentences produced in all sessions in the last column.

To test for stability of measures over time, we conducted three separate repeated-measures analyses of variance, one for each measure because the scale for each measure differed. The within-group variable was average score for the measure of interest averaged across the five sentence probes administered at each time point (Time 1, Time 2, and Time 3). For example, the mean SAS at Time 1 was the average score derived from the five sentences administered at that time point. For TNW, the model was significant, $F(2,28) = 4.78$, $p = .02$, partial $\eta^2 = .25$, and follow-up pairwise comparisons indicated that the TNW at Time 2 was significantly higher than at both Time 1 and Time 3, $p = .02$, with no difference between Time 1 and Time 3, $p = .57$. This can be seen by the mean scores provided in Table 2. For SAS, the model was not significant, $F(2,28) = 0.52$, $p = .60$. For WAS, the model was also not significant, $F(2,28) = 0.97$, $p = .39$. The stability of the latter two measures over time can be seen in the means and standard deviations provided in Table 2.

Regarding how the experimental measures related to one another and to standardized measures (CELF-5, TOWL-4), Pearson correlations were evaluated across two patterns of relationships, first among experimental measures, and second between experimental and standardized measures (see Table 3). Because there was only one difference in the TNW across time points and because of the small sample size and purpose of the study, total average scores were used for the TNW, SAS, and WAS. These scores represent all participants' data averaged across all 15 sentences across time points (e.g., last column of Table 2). Among experimental measures, the WAS was significantly related to both the TNW and the SAS; the direction was positive and the magnitude moderate for TNW and large for SAS. The SAS and the TNW were not significantly correlated. Between experimental and standardized measures, there

Table 3. Pearson correlations among measures

	1	2	3	4	5	6
1. Total number of words	–					
2. Sentence accuracy score	0.36	–				
3. Word accuracy score	0.58*	0.77**	–			
4. CELF-5; Formulated Sentences SS	0.40	0.83**	0.58*	–		
5. CELF-5; Recalling Sentences SS	0.26	0.59*	0.50	0.62*	–	
6. TOWL-4; Sentence Combining SS	0.31	0.47	0.30	0.56*	0.36	–

Note. CELF-5 = Clinical Evaluation of Language Fundamentals, Fifth Edition (Wiig et al., 2013); SS = Standard score; TOWL-4 = Test of Written Language, Fourth Edition (Hammill & Larsen, 2009).
*$p < .05$; **$p < .01$.

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

were three significant correlations. The SAS was significantly and positively related to both CELF-5 subtests and was interpreted as large for Formulated Sentences and moderate for Recalling Sentences. The WAS was also moderately positively related to the Formulated Sentences subtest. No experimental measures were observed to be significantly related to the TOWL-4 Sentence Combining subtest. Finally, it should be noted that there were two significant and moderately positive correlations observed among standardized tests; these were between the two subtests of the CELF-5 as well as between the TOWL-4 Sentence Combining subtest and the CELF-5 Formulated Sentences subtest.

As a follow-up to these findings, we wanted to evaluate the utility of the words selected, so we rank ordered the 15 stimulus words by average score across words for each of the three experimental measures: TNW, SAS, and WAS. Table 4 includes the list of stimuli words rank ordered from highest to lowest for each of the three experimental measures. Words are coded such that words administered at Time 1 are in bold, Time 2 in italics, and Time 3 in regular type. As can be seen from the patterns of rank ordering of words in Table 4, there was no discernable pattern for any time point or word for any of the measures. It should be noted that the word *since* resulted in the lowest SAS and WAS as the word is often misused to represent the word *because* and we were strict with our scoring of the word as a temporal adverb.

Research question 3

To address the third research question about end user feedback and in line with the pilot nature of the study, we asked the caregivers about their experiences with telepractice. As can be seen in the data presented in Table 1, caregivers indicated that the sessions were easy to navigate as the platform was familiar to both them and their children, given their experiences participating in remote schooling during the pandemic. That is, 93% responded "strongly agree" and 7% rated themselves "neutral" on interview items 1 and 4.

Table 5 includes selected responses to the open-ended question representing different types of feedback received. In response to

Table 4. Rank order of words (with mean scores) for each measure from highest to lowest across three time points (Time 1 in bold, Time 2 in italics, and Time 3 in regular type)

Total Number of Words		Sentence Accuracy Score		Word Accuracy Score	
Word	**Mean Score**	**Word**	**Mean Score**	**Word**	**Mean Score**
yet	*11.53*	*yet*	*0.91*	**because**	**2.00**
so	10.93	*gave*	*0.83*	*gave*	*1.80*
however	*10.80*	**because**	**0.81**	went	1.80
airplane	*10.73*	**jump**	**0.80**	for	1.73
but	**10.60**	*airplane*	*0.79*	**dinosaur**	**1.67**
also	**10.47**	*however*	*0.79*	so	1.67
went	10.20	so	0.79	instead	1.67
since	*10.00*	for	0.78	*however*	*1.53*
because	**9.47**	went	0.76	*airplane*	*1.47*
for	9.40	instead	0.74	**also**	**1.40**
gave	*9.07*	furthermore	0.72	*yet*	*1.40*
instead	8.73	**dinosaur**	**0.72**	**jump**	**1.27**
furthermore	8.53	**also**	**0.70**	furthermore	1.00
jump	**8.40**	**but**	**0.69**	**but**	**0.93**
dinosaur	**7.73**	*since*	*0.58*	*since*	*0.87*

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Table 5. Selected open-ended responses from caregivers

Interview Question	Response
Please share any comments you have about the remote writing research project that you would like us to know. We are particularly interested in ways you think we could have made the task more user-friendly. By user, we mean both you and your child who participated in the project.	"I liked to see the whole process of the study. When he brings things home from school, I only see the end product. I don't hear the instructions or see the process. It was cool to see the whole process." "It wasn't too time intensive. He could independently use the platform; he had used virtual platforms before because school was remote. It was easy to participate." "No feedback on improvements. In person, you would give kids brain breaks, but the memes were a creative solution to giving breaks since we weren't in person. This was a simple and straightforward process." "We were very familiar with [the virtual platform], so that worked very well for us to meet." "For my son, it would've been easier not online. He enjoys interaction with people. He did better than I thought he would. It wasn't difficult for him to do. He tried hard for the students." "Sometimes for the child it felt like more schoolwork to do. Sometimes he would get distracted by things around him at home, and in the future, it would be great if this could take place in person." "It would be beneficial for us [the researchers and parents] to hold up the writing sample and take a screenshot of the writing at the end of the session, rather than having the parents scan and e-mail it."

the open feedback/comments portion of the follow-up interview, two caregivers suggested that their children may have performed differently if the sessions were conducted in person. For example, one caregiver indicated that she thought the child became distracted by other things in his environment, although this was not observed by the graduate students and did not seem to impact performance. Suggestions for study enhancements included the use of video modeling to explain procedures to participants rather than explaining directions in verbal format only. Additional suggestions were related to submission methods for the writing samples, including holding the sentences up to the camera for the graduate students to take a screenshot of the child's sentences. This procedure would eliminate the need for the caregivers to spend time sending writing samples following the sessions.

DISCUSSION

The purpose of this study was to describe the development and demonstrate the feasibility of a sentence-writing probe assessment task administered through videoconferencing software and contribute to a necessary body of research on the utility of telepractice for writing assessment. The sample included students who were identified as struggling or typical writers by their caregivers, allowing for demonstration of the usability of the task by a range of writers. Findings from this study support the initial feasibility of virtual assessment using the sentence-writing probes to assess word- and sentence-level writing in intermediate-grade students.

The sentence-writing probe procedures were established by aligning stimuli selection and scaffolded elicitation procedures with current research (e.g., Arfé & Pizzocaro,

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

2016; Heilmann & Malone, 2014; Koutsoftas & Petersen, 2017; Ritchey et al., 2016). In line with Ritchey et al. (2016), the task provided visual, verbal, and written scaffolds by including picture stimuli, verbal rehearsal opportunities, and written key words needed to demonstrate writing skills at the word and sentence levels. Visual scaffolds included age-appropriate pictures that provided background knowledge and visual stimuli for generating ideas to write. Written scaffolds were twofold and included a single-clause sentence and the target word written on the slide alongside the picture. In this way, students could simply generate a complex sentence from what they were provided or create a new novel sentence. In either case, written words related to the picture were provided in the form of a sentence, which reduced transcription constraints related to spelling and grammar (Graham, 1990) and supported students in their generation of sentences. Verbal scaffolds included the examiner reading the written text and providing the student an opportunity for verbal rehearsal of what they planned on writing. This allowed for text generation independent of transcription. After the child verbalized, they wrote the sentence and this provided an opportunity to evaluate both text generation and transcription skills.

The pattern of scores obtained across measures (TNW, SAS, and WAS) over time and how these were related to one another and standardized tests provide evidence of initial feasibility for use in subsequent studies. The pattern of scores over time did not change much, which was what we expected, given that the task was not associated with an intervention and because of the short duration between time points (i.e., 2 weeks between probes). There was one exception whereby there were significantly more total words produced at Time 2 by an average of about one word. There was no discernable pattern observed in the rank ordering of words by time point across measures, suggesting parity between the stimuli presented at each time point. The rank ordering of our results using mean SAS and WAS further supported that there was no pattern or advantage for words, except the words *since* and *furthermore*, which proved to be the most challenging for students in this sample. For the word *since*, students did use the word as a causal connective; however, we were strict in our scoring of it as a temporal adverb. For the word *furthermore*, participants at times either did not complete their sentence or had difficulty using the word as an additive conjunction. Two example sentences from the participants include "The dogs are outside furthermore," and "The dog on the right has a furthermore part of the stick." It is possible that if these words were targeted as part of an intervention, students would show gains over time. Future studies are necessary to see whether the response pattern changes when the task is associated with a specific intervention or whether the words should be randomized across participants instead of across time points.

Measures obtained from averages of scores from all 15 sentences used across probes proved to be indicators of sentence-level written language, as these were related with one another and with standardized measures of language. The TNW is considered a measure of writing productivity and was not related to any standardized measure; however, it was related to the WAS. The WAS indicated correct use of a word to build a sentence and is considered a word-level measure, whereas the SAS is considered a sentence-level measure of writing accuracy and mechanics. Because TNW accounts for number of words and WAS is a word-level measure, these were found to be related. Both measures were significantly related to standardized measures of oral language though not written language. The Sentence Combining subtest of the TOWL-4 (Hammill & Larsen, 2009) was, in fact, not related to any experimental measure. More specifically, sentence-level measures of oral language from the CELF-5 (Wiig et al., 2013) were related to both measures of writing from the sentence-writing probes but not to the standardized

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

measure of writing. Our conjecture is that, because the sentence-writing probe procedure included a verbal rehearsal component, it was related to the oral language measures on the CELF. The TOWL-4 Sentence Combining subtest does not evaluate oral language components. It is important to note that the sentence-writing probe task was developed on the basis of the Formulated Sentences subtest of the CELF-5, so it makes sense that there were significant relationships between that subtest and sentence-writing probe measures. The differences between the tasks were that in the sentence-writing probe, students were (a) provided with written words and sentences and (b) provided time to plan their response and verbally rehearse before writing the sentence. Future research should evaluate the use of these measures in a larger sample where validity and reliability of the measures can be evaluated.

The use of telepractice to administer the sentence-writing probe task showed benefits of increased attendance and convenience for families. For example, families residing a distance from the university did not have to travel to participate in the study. Many of the parents/caregivers reported that the study was easy to participate in, given their familiarity with videoconferencing platforms used during the COVID-19 pandemic. A few caregivers commented that their child may have performed differently if the sessions were held in person. For future telepractice, researchers and clinicians need to consider student characteristics and caregiver preference when deciding modality of service as this aligns with evidence-based practice. These findings suggest that the flexibility of scheduling virtual assessment procedures shows promise as a way to evaluate progress in the context of intervention.

Adjustment to the procedures for gathering the participants' writing samples from the caregivers may have prevented the attrition of two participants due to missing data points. One suggestion from a caregiver was to hold the sentences up to the camera for the graduate students to take a screenshot of the written work. We did attempt this; however, the image was not as clear as when provided with a photograph or scan using a smartphone. Regardless, the clarity of the written sentences using a provided image was not always acceptable; however, during administration, examiners recorded verbal recitations by the students of what they wrote. These data helped clarify spelling and handwriting. It is important to have students read their final written work and record this response for subsequent analyses. Future studies should ensure that transmission of data by parents/caregivers is addressed as part of the study design. In some cases, parents sent images immediately after the session, others were delayed, and only two parents lost data during the study.

The impetus for this study was to develop a sentence-level measure of writing for use in subsequent interventions and research studies that employ single-case research designs. Single-case research designs are often used for writing intervention studies (e.g., Rogers & Graham, 2008) to demonstrate the feasibility or early efficacy of an intervention or for testing interventions with populations where heterogenous groups of participants are a challenge to attain (Fey & Finestack, 2009; Kratochwill & Levin, 2014). Single-case research designs require that measures of the target behavior (i.e., the dependent variable) are provided with regularity (Kratochwill et al., 2013, 2023). The sentence-writing probes presented in this study provide a potential solution as they can be integrated into single-case research designs as a dependent measure of word- and sentence-level writing. Although our findings generally did not show significant differences between time points, changes may be seen in practice or research when students are participating in an intervention.

According to Kratochwill et al. (2013), single-case designs meet standards, as opposed to meet standards with reservation, if they include a minimum number of phases and data points across differing designs (e.g., ABAB design, multiple-baseline design). To

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

meet standards, a range of four to six phases of observation (or probes) are necessary and can span the baseline and maintenance phases. Within each phase, the required number of data points is generally a minimum of 5 (Kratochwill et al., 2023). This sentence-writing probe task aligns with these standards in that each probe included five sentences (or data points); however, in this study we did not examine the utility of the probes within an intervention context. Future studies will require the inclusion of additional probes across phases to align with the What Works Clearinghouse single-case design standards (Kratochwill et al., 2013).

Finally, in considering the domain of writing for assessment, it is important to consider the role of oral language in this process. This is especially so for students with special education needs such as developmental language disorders, language-based learning disabilities, or autism. It may be the case that limitations in oral language constrain text generation processes associated with writing. This sentence-writing probe task required students to think, say, and write, and provided multiple scaffolds for formulating sentences, including verbal, visual, and written ones that support the writing process (Ritchey et al., 2016). For example, students who can generate text orally by engaging in verbal rehearsal but do not produce complete sentences might be constrained by transcription limitations. The sentence-writing probes allowed for observation of this so that clinicians and researchers might better understand the individual differences students exhibit when writing sentences, and how interventions can support improvements in these skills.

Limitations and future directions

This was a feasibility study with the goal of demonstrating proof of concept for a virtual assessment probe task of sentence-level writing that could be further developed and integrated into research designs, specifically single-case designs. Although we achieved the demonstration of proof of concept, important future considerations for this work have been identified through limitations in the current study. These data were collected from one geographic region with a very small sample, so a larger and more diverse sample is needed for future study, especially with regard to rural and urban settings. One challenge was the process of collecting participants' written sentences following each session. Other methods for sharing writing samples should be explored to enhance this process, while keeping in mind the quality of the copy of the sample. Another limitation of this feasibility study was that end user feedback was gathered only from the caregivers. In future studies, feedback should be gathered from the child participants to make adjustments to the probe task and as a qualitative measure of level of difficulty of target words. The selection of words included in the probes for the current study was based on prior reports of frequency of use of cohesive ties, which were then randomized across time points. In doing so, we limited the types of complex sentences to only subordinating conjunctions, yet intermediate-grade students are expected to produce a variety of complex sentences. Although we do suggest that future research retain the use of nouns and verbs in probes to support minimal written output from participants, future studies should also consider the types of words included in the probes and how these relate to developmental expectations and alignment with an associated intervention.

CONCLUSION

This study demonstrated the initial feasibility and proof of concept of this sentence-writing probe task for administration via telepractice. Findings from this study show promise for the usability of this task with intermediate-grade students, with clear directions for future research. The remote learning format was particularly beneficial for families of children in our study who lived in rural areas as they would not have to encumber the expense and time of traveling to a school

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

or university clinic in a larger metropolitan area to receive services or participate in research studies. With further development, this sentence-writing assessment task can be used to document the functional relationship between an intervention and writing outcomes at the word and sentence levels for both research and clinical purposes.

REFERENCES

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, *102*(2), 281–298. https://doi.org/10.1037/a0019318

Arfé, B., & Pizzocaro, E. (2016). Sentence generation in children with and without problems of written expression. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and spoken language development across the lifespan: Essays in honour of Liliana Tolchinsky* (pp. 327–344). Springer. https://doi.org/10.1007/978-3-319-21136-7_19

Berninger, V. W. (2009). Highlights of programmatic, interdisciplinary research on writing. *Learning Disabilities Research & Practice*, *24*(2), 69–80. https://doi.org/10.1111/j.1540-5826.2009.00281.x

Berninger, V. W., Abbott, R. D., Swanson, H. L., Lovitt, D., Trivedi, P., Lin, S. J., Gould, L., Youngstrom, M., Shimada, S., & Amtmann, D. (2010). Relationship of word- and sentence-level working memory to reading and writing in second, fourth, and sixth grade. *Language, Speech, and Hearing Services in Schools*, *41*(2), 179–193. https://doi.org/10.1044/0161-1461(2009/08-0002)

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). Guilford Press.

Berninger, V., Whitaker, D., Feng, Y., Swanson, H. L., & Abbott, R. D. (1996). Assessment of planning, translating, and revising in junior high writers. *Journal of School Psychology*, *34*(1), 23–52. https://doi.org/10.1016/0022-4405(95)00024-0

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Dockrell, J. E., Ricketts, J., Charman, T., & Lindsay, G. (2014). Exploring writing products in students with language impairments and autism spectrum disorder. *Learning and Instruction*, *32*, 81–90. https://doi.org/10.1016/j.learninstruc.2014.01.008

Fey, M. E., & Finestack, L. H. (2009). Research and development in child language intervention: A five-phase model. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513–531). Psychology Press.

García, J. N., & Fidalgo, R. (2008). The orchestration of writing processes and writing products: A comparison of 6th grade students with and without learning disabilities. *Learning Disabilities: A Contemporary Journal*, *6*(2), 77–98.

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology*, *82*(4), 781–791. https://doi.org/10.1037/0022-0663.82.4.781

Hammill, D. D., & Larsen, S. C. (2009). *The test of written language—fourth edition*. ProEd.

Heilmann, J., & Malone, T. O. (2014). The rules of the game: Properties of a database of expository language samples. *Language, Speech, & Hearing Services in Schools*, *45*(4), 277–290. https://doi.org/10.1044/2014_LSHSS-13-0050

Koutsoftas, A. D. (2018). Writing-process products of fourth- and sixth-grade children: A descriptive study. *The Elementary School Journal*, *118*, 632–653. https://doi.org/10.1086/697510

Koutsoftas, A. D., & Petersen, V. (2017). Written cohesion in children with and without language learning disabilities. *International Journal of Language & Communication Disorders*, *52*(5), 612–625. https://doi.org/10.1111/1460-6984.12306

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., & Johnson, A. (2023). Single-case intervention research design standards: Additional proposed upgrades and future directions. *Journal of School Psychology*, *97*, 192–216. https://doi.org/10.1016/j.jsp.2022.12.002

Kratochwill, T. R., & Levin, J. R. (2014). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, *15*(2), 124–144. https://doi.org/10.1037/a0017736

McCrimmon, A.W., & Climie, E.A. (2011). Test review: Test of written language—fourth edition. *Journal of Psychoeducational Assessment*, *29*(6), 592–596. https://doi.org/10.1177/0734282911406646

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psy-*

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

chology Review, *8*(3), 299–325. http://www.jstor.org/stable/23359419

McMaster, K. L., Du, X., & Pétursdóttir, A. L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, *42*(1), 41–60. https://doi.org/10.1177/0022219408326212

McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children*, *77*(2), 185–206. https://doi.org/10.1177/001440291107700203

Olswang, L. B. (1998). Treatment efficacy research. In C. M. Frattali (Ed.), *Measuring outcomes in speech-language pathology* (pp. 134–150). Thieme.

Platt, A., & D'Anna, R. (2022). *Pilot and feasibility studies*. Research Design and Analysis Core Guide. Duke Global Health Institute. https://sites.globalhealth.duke.edu/rdac/wp-content/uploads/sites/27/2022/06/Core_Guide_Pilot_and_Feasibility-Studies_06_17_22.pdf

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*, *17*(2), 107–120. https://doi.org/10.1044/1058-0360(2008/012)

Ritchey, K. D., & Coker, D. L. (2013). An investigation of the validity and utility of two curriculum-based measurement writing tasks. *Reading & Writing Quarterly*, *29*, 89–119. https://doi.org/10.1080/10573569.2013.741957

Ritchey, K. D., McMaster, K. L., Al Otaiba, S., Puranik, C., Kim, Y.-S. G., Parker, D. C., & Ortiz, M. B. (2016). Indicators of fluent writing in beginning writers. In K. Cummings & Y. Petscher (Eds.), *The fluency construct* (pp. 21–66). Springer. https://doi.org/10.1007/978-1-4939-2803-3_2

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology*, *100*(4), 879–906. https://doi.org/10.1037/0022-0663.100.4.879

Vanderberg, R., & Swanson, H. L. (2007). Which components of working memory are important in the writing process? *Reading and Writing*, *20*, 721–752 https://doi.org/10.1007/s11145-006-9046-6

Whitaker, D. P., Berninger, V. W., Johnston, J. G., & Swanson, H. L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences*, *6*(1), 107–130. https://doi.org/10.1016/1041-6080(94)90016-7

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals—fifth edition*. Pearson.

Appendix. Target words by probe in order of administration

Time 1
dinosaur (noun)
jump (verb)
but (adversative/contrastive)
because (causal)
also (additive)
Time 2
airplane (noun)
gave (verb)
since (temporal)
however (adversative/contrastive)
yet (adversative/contrastive)
Time 3
went (verb)
furthermore (additive)
so (causal)
for (causal)
instead (adversative/contrastive)

Copyright © 2023 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.