Internet-Based Telehealth Assessment of Language Using the CELF–4

Monique C. Waite
Deborah G. Theodoros
Trevor G. Russell
The University of Queensland, Brisbane, Australia
Louise M. Cahill
Royal Children's Hospital, Brisbane, Australia

Expressive and receptive language disorders account for the majority of new referrals to pediatric speech-language pathology services (Broomfield & Dodd, 2004). Prevalence figures for language impairment in the primary school-age population have ranged from 3.1% to 23.9% (Blum-Harasty & Rosenthal, 1992). Recent Australian data suggest an even higher prevalence of 32.5% (either language impairment or comorbid speech and language impairment) for preparatory students (Jessup, Ward, Cahill, & Keating, 2008). Children with language impairment at a young age face the risk of continuing difficulties in language and related tasks in their academic life and in adulthood (e.g., Bishop & Adams, 1990; Clegg, Hollis, Mawhood, & Rutter, 2005; Johnson et al., 1999; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998; Young et al., 2002). Language impairment is also associated with social, emotional, and behavioral problems, which may persist into later life (e.g., Baker & Cantwell, 1987; Beitchman et al., 1996; Cantwell & Baker, 1987).

Given the lifelong impact that a language disorder in childhood can have on an individual, early identification and thorough and specific assessment and treatment are imperative. However, access to speech-language pathology services may be difficult for many children and their families, particularly those residing in rural and remote areas (American Speech-Language-Hearing Association [ASHA], 1985; Chezik, Pratt, Stewart, & Deal, 1989; O'Callaghan, McAllister, & Wilson, 2005; Wilson, Lincoln, & Onslow, 2002). A survey of caregivers of children in rural and remote New South Wales found that the majority of families experienced barriers to accessing speech-language pathology services, with the lack of speech-language pathologists (SLPs) in these areas identified as the primary concern by the majority (85%) of the informants (O'Callaghan et al., 2005). Labor force data support the issue of

ABSTRACT: **Purpose:** Telehealth has the potential to improve children's access to speech-language pathology services. Validation of telehealth applications, including the assessment of childhood language disorders, is necessary for telehealth to become an accepted alternative mode of service provision. The aim of this study was to validate an Internet-based telehealth system for assessing childhood language disorders.
Method: Twenty-five children ages 5 to 9 years were assessed using the core language subtests of the Clinical Evaluation of Language Fundamentals—4th Edition (CELF–4; Semel, Wiig, & Secord, 2003). Each participant was simultaneously assessed online and face-to-face (FTF). Assessments were administered by either an online or an FTF speech-language pathologist (SLP), but were simultaneously rated by both SLPs.
Results: No significant difference was found between the online and FTF total raw scores and scaled scores for each subtest. Weighted kappas revealed very good agreement on the individual items, total raw scores, scaled scores, core language score, and severity level. Intra- and interrater reliability were determined for a sample of online ratings, with intraclass correlation analysis revealing very good agreement on all measures.
Conclusion: The results of this study support the validity and reliability of scoring the core language subtests of the CELF–4 via telehealth.

KEY WORDS: Internet-based, telehealth, telepractice, standardized language assessment

reduced employment of SLPs in rural and remote areas of Australia (Fitzgerald, Hornsby, & Hudson, 2000; Lambier & Atherton, 2003). In a 2003 survey of members of the Speech Pathology Association of Australia, it was reported that only 3.9% of Australia's SLPs were primarily employed in moderately accessible, remote, or very remote areas, whereas 94.7% were employed in highly accessible or accessible areas (Lambier & Atherton, 2003). The low proportion of SLPs working in rural and remote areas of Australia is due to limited availability of health services (Humphreys, Hegney, Lipscombe, Gregory, & Chater, 2002) and difficulties recruiting and retaining SLPs (Fitzgerald et al., 2000; National Rural Health Alliance, 2004). Considering that 34% of Australians live in regional or remote areas (Australian Bureau of Statistics, 2003), the lack of speech-language pathology services in these areas is a concern. The challenges of providing equitable services to rural and remote regions are not unique to Australia (Pickering et al., 1998). In 2005, 19% of the U.S. population lived in rural areas (Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, 2009). This substantial proportion of the population is likely to be underserved in terms of speech-language pathology services.

An Australian study (O'Callaghan et al., 2005) has identified additional barriers to accessing pediatric speech-language pathology services in rural and remote areas, such as long distances, costs involved in accessing services, long waiting lists, reduced service choices, lack of public transport, and reduced awareness of available services. Similar issues have been raised in the United States (ASHA, 1985). However, difficulties accessing speech-language pathology services are not restricted to people living in rural areas. For example, if parents are required to work, have other children, or are facing financial difficulties, regular attendance at a speech-language pathology clinic may be difficult. Increasing demand for speech-language pathology services has an impact on metropolitan as well as rural areas. Issues such as long waiting lists (Victorian Department of Human Services, 2004) and difficulty recruiting experienced staff (Iacano, Johnson, Humphreys, & McAllister, 2007; Victorian Department of Human Services, 2004) have been noted as a concern in metropolitan areas of Australia. In the United States, a shortage of SLPs exists in the health (ASHA, 2007) and education (ASHA, 2008) sectors in metropolitan, suburban, and rural areas.

Telehealth—the provision of health services remotely using information technology and telecommunications (also called telepractice)—offers a potentially viable solution to the difficulties in providing speech-language pathology services, as either an alternative or an adjunct service delivery model. In 2005, an ASHA working group recognized telepractice as an appropriate mode of speech-language pathology service delivery. Successful telepractice programs have since emerged, such as INTEGRIS health's speech teletherapy program, which provides services to schools in rural Oklahoma (Forducey, 2006; Scheideman-Miller et al., 2002). Despite the reported success of such programs, there is a lack of published empirical studies of telehealth applications in speech-language pathology (Hill & Theodoros, 2002). Research into telehealth assessment of pediatric communication disorders is particularly lacking, and the small numbers of investigations that have been conducted are limited to single-case (Savard, Borstad, Tkachuck, Lauderdale, & Conroy, 2003) or small-scale pilot studies (Cole, Martin, Moody, & Miller, 1986; Fairweather, Parkin, & Rozsa, 2004; Waite, Cahill, Theodoros, Busuttin, & Russell, 2006).

A pilot study in 1986 investigated the telehealth assessment of school-age children's speech and language (Cole et al., 1986). This study used an audio teleconferencing system with a microphone and loudspeaker that operated via a telephone landline. Five children were assessed simultaneously using the Renfrew Action Picture Test (RAPT; Renfrew, 1971) and the Fisher-Logemann Test of Articulation Competence (Fisher & Logemann, 1971) by both an SLP who was located in the same room as the child and a remote SLP who directed the session. Cole et al. (1986) reported that there were no differences between the two SLPs' ratings on the RAPT, thus demonstrating the adequacy of an audio-only system for this assessment of expressive language. However, the study did not include any measure of receptive language. Commonly, receptive tasks require the child to point to items in response to the clinician's direction. This activity would be difficult to assess using the audio-only system, as the remote clinician would be unable to see the child's responses on such tasks. The lack of video did cause a problem for the articulation assessment, the Fisher-Logemann Test of Articulation Competence, as in one case, the remote clinician could not perceive an interdental lisp because she was unable to view the child's tongue placement (Cole et al., 1986). Despite these limitations of the telehealth system, the overall positive results did demonstrate the potential of such a simple system for delivering certain speech-language pathology services. However, the results were only preliminary, and no larger scale studies using this system have been reported.

Videoconferencing may offer greater potential for the evaluation of speech and language across the Internet. In 2003, Savard et al. reported the real-time assessment of the language, cognition, and oral motor control of a 9-year-old boy with hemorrhagic dengue fever via a three-way videoconference. The videoconferencing system enabled the clinician to demonstrate each task for the oral motor assessment and then view the child's productions, which would not have been possible with the teleconferencing system used by Cole and colleagues (1986). From this session, the SLP was able to establish the child's major communication difficulties and provide recommendations for his care. The case study by Savard et al. exemplifies how successful evaluation of pediatric communication can be conducted via specialized videoconferencing equipment. The inclusion of video enabled the clinician to view the child's face so that she could assess his oral motor functioning. Although the SLP reported that the videoconferencing session seemed clinically effective, she indicated that there were no validated tools for evaluation and follow-up (Savard et al., 2003). The lack of validated assessment tools remains a barrier for the widespread use of telehealth in the field of speech-language pathology.

A more recent pilot study (Fairweather et al., 2004) investigated the assessment of language, speech, and oral motor functioning of 13 children ages 6 to 14 years, using both informal and standardized tests. The assessments were conducted via interactive videoconferencing over an integrated services digital network connection. As with Cole and colleague's (1986) study, all the assessments were conducted by a remote SLP but were simultaneously rated by the remote SLP and a second SLP who was located in the same room as the child. The on-site clinician maneuvered the equipment and stimulus materials following the remote SLP's directions. All the children were assessed with two standardized assessments, the Goldman-Fristoe Test of Articulation—2nd Edition (Goldman & Fristoe) and the Clinical Evaluation of Language Fundamentals—3rd Edition (CELF–3; Semel, Wiig, & Secord, 1995), as well as an informal conversational speech sample analysis and oromotor assessment.

The overall results of this pilot study were positive, as there was 69% and 92% total agreement by the two SLPs on the severity of the children's articulation disorders and language disorders, respectively (Fairweather et al., 2004). For ratings made on the CELF–3, there was complete agreement for the individual subtests for 69% of the participants. Interrater discrepancies were attributed to subjective scoring differences and difficulty differentiating words spoken by the participants. It was not stated whether this difficulty was due to the audio quality of the videoconference or reduced intelligibility of the participants. High overall agreement for the various observations and severity judgement was determined. Several discrepancies in the clinicians' judgment of individual errors in the articulation test were found to occur in participants with severe articulation disorders and on certain classes of sounds (Fairweather et al., 2004). The researchers therefore felt that the protocol was more useful for assessing mild speech difficulties rather than moderate to severe disorders. Generally high levels of agreement between the two SLPs for the oromotor assessment were reported. However, the researchers commented that videoconferencing assessments were limited due to problems in obtaining sufficient viewing angles and poor lighting making the assessment of certain structures and functions (e.g., velopharyngeal system) difficult. The authors believed that telehealth has the most potential for assessing children and adolescents ages 6 to 16 years, as modifications such as including more tangible reinforcement and changing furniture placement would be required for younger children (Fairweather et al., 2004).

Although videoconferencing technology provides the capacity for a wider range of pediatric communication assessment to be conducted via telehealth, the technologies used in the two previous studies (Fairweather et al., 2004; Savard et al., 2003) had some limitations. First, a second person was required to manipulate the stimulus materials at the participant's site. In the clinical setting, this would require extra time training a parent or aide as well as transporting equipment to the remote site. Second, as all ratings were carried out during the videoconference, it is possible that limitations in the audio and video quality, such as audio breakup and reduced image quality, contributed to the interrater differences. Third, this type of videoconferencing equipment may be problematic due to the costs of installation and maintenance and the need for the client to travel to a room-based videoconferencing suite (Wilson, McAllister, Atkinson, & Sefton, 2006).

In contrast, a PC-based, desktop videoconferencing system that operates over an Internet Protocol (IP) connection may have greater utility as a telehealth system. This type of system would be less costly and would use the equipment and telecommunication infrastructure that are already available to the client. A PC-system would enable the computerization of stimuli so that test materials would not need to be transported to the child. Digitally recorded audio or visual stimuli could be delivered systematically, achieving greater consistency in administration and minimizing examiner bias (Haaf, Duncan, Skarakis-Doyle, Carew, & Kapitan, 1999). This type of system would also be able to record and store the child's responses. Videoconferencing at low bandwidths may be inadequate to judge certain speech errors (Wilson et al., 2006) and is associated with issues of audio and visual breakup (Theodoros, Hill, Russell, Ward, & Wootton, 2008). Making quality recordings of client responses would overcome these limitations, enabling more accurate scoring.

A number of studies have investigated the use of computer-based telehealth systems to assess adult communication disorders (Brennan, Georgeadis, Baron, & Barker, 2004; Hill, Theodoros, Russell, & Ward, 2008; Hill et al., 2006; Lemaire, Boudrias, & Greene, 2001; Theodoros et al., 2008; Ward et al., 2007). To date, there have been no investigations of the use of such a system for assessing children's language. However, the potential of computer-based language assessment has been exemplified by an investigation by Haaf et al. (1999). This study aimed to validate computer administration of a standardized language assessment, the Peabody Picture Vocabulary Test—Revised (PPVT–R; Dunn & Dunn, 1981). Seventy-two normally developing children, ages between 4;0 (years;months) and 8;11, were randomly assigned to one of three groups. The first group followed the standard assessment procedure. In the other two groups, scanned stimulus images and digitally recorded verbal stimuli were presented on a computer. In one group, the child used a trackball to move the cursor and select a picture. In another group (automated scanning), the computer automatically highlighted each picture for 2 s and the participant responded by pushing the trackball button. The authors reported no statistically significant differences in performance across the three response conditions, thus supporting the study hypothesis that the computerized test administration was equivalent to the standard administration (Haaf et al., 1999). Such a computerized assessment could be used for telehealth purposes, as the SLP does not need to be present to administer it. Computer adaptation of language assessments has the potential to be used in conjunction with videoconferencing, which would also enable the SLP to interact with the child and record observational data.

One study has investigated the assessment of children via PC-based videoconferencing (Waite et al., 2006). In this study, six children, ages between 4 and 9 years, with identified speech impairment were assessed via desktop videoconferencing, operating through a 128-kilobits/sec (Kbs) Internet link. Each child was simultaneously assessed by a remote (online) and a face-to-face (FTF) SLP on single-word articulation, oral-motor structure and function, and conversational speech intelligibility. The online system enabled demonstration videos and stimulus pictures to be presented on the participant's computer screen. High-quality audio and video clips of each participant's responses could be collected, stored, and transmitted to the online SLP. Comparisons of the ratings made by the online and FTF clinicians yielded positive results, with high levels of overall agreement for ratings of single-word articulation (92%), oral-motor tasks (91%), and speech intelligibility (100%) (Waite et al., 2006). High intra- and interrater agreement were achieved for online ratings of most measures. The positive results suggest that an Internet-based protocol has potential for the assessment of pediatric speech disorders. The desktop system enabled the flexibility to display stimulus materials on the participant's computer and to video- and audio-record participant responses. It has been suggested that the flexibility offered by computer-based telehealth may enable the delivery of any established speech-language pathology practice, including delivery of standardized assessments, to be conducted online (Hill et al., 2006). However, the findings of the study by Waite and colleagues (2006) were only preliminary, and further investigation of other client groups and use of larger sample sizes is needed.

Although previous studies of telehealth assessment of childhood language disorders have been positive, further research into different telehealth technology is required to overcome limitations in telehealth methods previously investigated, particularly those associated with audio and visual quality, and logistical issues such as transporting physical materials to the client. The aim of this study was to examine the validity and reliability of an Internet-based telehealth system for assessing childhood language disorders on the core components of a standardized assessment tool. The research sought to determine the technical feasibility of the Internet-based telehealth system in order to answer the following research questions:

- Is the scoring of the core components of a standardized language assessment via telehealth equivalent to scoring the assessment in a traditional FTF environment?
- Does scoring via telehealth demonstrate adequate intra- and interrater reliability?

METHOD

Participants

Twenty-five children between 5 and 9 years of age (M_{age} = 7;2, *SD* = 1;3) participated in the study. There were four 5-year-olds, eight 6-year-olds, three 7-year-olds, and ten 8-year-olds. Approximately two thirds (17) of the participants were male and one third (8) were female. The children either had been previously diagnosed with language impairment by an SLP (*n* = 12; severity ranging from mild to significant) or were identified with difficulties in language by a parent or teacher but had not been formally assessed by an SLP (*n* = 13). Participants were excluded from the study if they had a significant hearing impairment, uncorrected visual impairment, or significant neurological or cognitive impairment, or if English was not their primary language. No previous experience with the Internet or computers was necessary as the child was not required to operate the equipment. Consent was obtained in writing from a parent or guardian and verbally from the child before participation commenced.

SLPs

A total of three qualified SLPs participated in the assessment sessions. The principal investigator of the study took part in all 25 assessments. The other two SLPs were selected by convenience, one taking part in 16 assessments and the other taking part in nine. The SLPs had a range of experience in the assessment and treatment of pediatric language disorders (2 to 10 years after graduation) and were experienced in the administration and scoring of the standardized assessment. Each SLP was trained in the setup and operation of the telehealth system and administration of the online assessment until she was able to demonstrate her proficiency in administering a simulated assessment with another SLP in the role of the participant. The training took approximately 3 hr over several sessions. In addition, each clinician was provided with a manual detailing the steps in the online administration and was able to refer to this during the assessments as needed.

Procedure

Each assessment was conducted between two rooms within the same building at the University of Queensland. All assessments used a telehealth system that provided real-time videoconferencing through a 128-Kbs Internet link (Hill et al., 2006). Each participant was allocated to one of two administration conditions—an online-led or an FTF-led assessment. Two of the three SLPs took part in each online-led and FTF-led assessment session. An FTF SLP was located with the child at the participant site. An online SLP was located in the second room. For each session, either the online SLP or the FTF SLP administered and scored the assessment while the other SLP simultaneously rated the assessment as an observer. As slightly different procedures were used for scoring by each SLP in the two administration conditions, there were a total of four scoring conditions: online scoring in the online-led assessment, online scoring in the FTF-led assessment, FTF scoring in the online-led assessment, and FTF scoring in the FTF-led assessment.

The simultaneous scoring by both clinicians allowed for comparison of the ratings obtained in each environment while eliminating the test–retest or test-learning effects that may have occurred if the participant was assessed online and FTF on two separate occasions. The FTF-led assessments were included to reduce any administration condition bias that may have occurred if the assessments were only led online. The authors have considered other research designs, such as a cross-over study, but extensive research experience in telehealth equivalence studies in other populations has revealed that test–retest effects are much greater than variability of performance in the online and FTF environments. The procedure used in the present investigation has been used in numerous published studies of telehealth assessments in speech-language pathology (Hill, Theodoros, Russell, & Ward, 2008; Hill et al., 2009; Theodoros et al., 2008; Waite et al., 2006).

For each assessment, the clinicians were randomly assigned their role, and the participant was randomly assigned to an administration type via a randomization Web site (www.randomization.com). In total, 13 online-led and 12 FTF-led assessments were conducted.

Online-led administration procedure. For the 13 online-led assessments, the online SLP directed each session and administered the assessment remotely via the telehealth system. The child was seated in front of a computer, and the online SLP established the videoconference, displayed stimuli, and recorded the child's responses via the system. The FTF SLP rated the assessment as an observer only. The FTF SLP transcribed and scored items during the assessment, but also recorded the entire session via a digital video camera. The digital video recordings were later converted to Windows Media Video files (encoded at 235 Kbs for video and 70 Kbs for audio) and saved onto a compact disc for completion of scoring of all subtests. Although the online SLP directed the entire session, some minimal interaction between the FTF SLP and the participant took place (e.g., the FTF SLP was occasionally required to reposition the child's headset, ensure that the child was seated correctly, and assist the child during short breaks between tasks). However, this interaction was directed by the online (i.e., leading) SLP. It was felt that this would be a realistic representation of a clinical telehealth assessment, where there would usually be another adult, such as a parent or teacher's aide, to assist with the session. See Figure 1 for a depiction of the online-led assessment setup at the participant site.

FTF-led administration procedure. This assessment was conducted according to standard clinical procedures. The participant and FTF SLP were positioned facing away from the computer system, and the assessment used the standard stimulus book. The

Figure 1. The telehealth system at the participant site during an online-led assessment.

Note. Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF®–4). Copyright © 2003 by NCS Pearson, Inc. Reproduced with permission. All rights reserved. "CELF" is a trademark, in the United States and/or other countries, of Pearson Education, Inc. or its affiliate(s).

FTF SLP directed the session and administered the assessment. The FTF SLP recorded the entire assessment via a digital video camera to enable completion of scoring. While the assessment was performed, the online SLP observed the session and recorded the child's responses via the telehealth system.

Assessment task. Each participant was assessed using the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF–4), Australian adaptation (Semel, Wiig, & Secord, 2003). For the purpose of the present study, permission to reproduce stimulus materials in a digitized format was obtained from the test publishers (Harcourt Assessment, Inc.). The four core language subtests for children ages 5 to 8 years (Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences) were administered. These subtests represent a set of tasks that are used to derive a measure of overall language ability—the core language composite score—which can be used as an indicator of a language impairment (Semel et al., 2003). We decided to administer only the core language subtests in order to include a selection of receptive and expressive tasks that could be administered in one assessment session but would provide sufficient information to identify the presence of a language impairment. We felt that the younger children might not be able to concentrate for a longer period in order to complete the optional subtests.

All subtests were administered in accordance with the testing procedures set out in the examiner's manual. In both administration conditions, the participant's spoken responses were audio recorded by the online SLP and video recorded by the FTF SLP. Both the online and FTF SLPs rated the responses during the assessment but were also required to check their ratings from the assessment recordings at a later time. For consistency, each rater was instructed to play back each recording a maximum of four times only. For each subtest, the total raw scores were converted to scaled scores, as per the procedure set out in the test manual (Semel et al., 2003). The scaled scores for each subtest were then totalled and were converted to a core language composite score; a severity level of language impairment was then determined (core language score $\geq$ 86 = normal, 78–85 = mild, 71–77 = moderate, $\leq$ 70 = significant; Semel et al., 2003).

System architecture. The assessments used a custom-built telehealth system, which has been detailed in a previous study (Hill et al., 2006). This system provided real-time videoconferencing through a 128-Kbs Internet link operating between two computers located at the participant and online clinician sites. This bandwidth was chosen to equal the minimum bandwidth in Queensland Health and Education Queensland facilities. At the participant site, two Web cameras (Logitech, QuickCam Pro 4000) mounted on a motorized base (Egletron, TrackerPod) were positioned above the computer monitor (see Figure 1). The motorized base enabled the online SLP to adjust the camera position remotely in order to ensure optimal viewing of the child during the online-led assessments.

For the FTF-led assessment, the Web cameras were positioned to view the participant pointing to items in the stimulus book. The first Web camera was used to view the participant during the videoconference; the second Web camera recorded video of the participant's responses. At the online SLP site, only one Web camera was positioned above the computer monitor for the purpose of the videoconference. Each participant wore a headphone/microphone headset to allow communication during the videoconference. The configuration of the child's headset was altered for either the online-led or FTF-led assessments. During the FTF-led assessment, only the microphone was operational so that the participant's responses could be heard by the online clinician, but he or she could not hear the online clinician.

In addition to videoconferencing, the telehealth system incorporated a number of features to enable the display of stimulus materials and recording of participant responses.

Stimulus display. This feature was used in the online-led assessment only. For all subtests except for Recalling Sentences, images were scanned from the stimulus book and saved in .jpg format. When displayed on the computer monitor, the images were identical in appearance to those in the stimulus book. During the online-led assessment, the online SLP selected the relevant image, which prompted the system to display it on the participant's monitor. The same image was also displayed on a second monitor at the online SLP's site. The system software enabled the online SLP to highlight a part of the image, which was then viewable by the participant.

For Recalling Sentences, which is a sentence repetition task, stimulus video recordings (.avi format) were made. These were included in order to ensure standard stimulus presentation that would be unaffected by breakdowns in image or audio quality that could occur during the videoconference. The online SLP selected each video and prompted the system to play it on the participant's computer. The auditory stimuli for the other subtests were delivered by the online SLP during the videoconference so that the assessment could be conducted more naturalistically. For instance, in the Word Structure subtest, it would have been difficult to coordinate pointing to part of the picture with an audio recording of the stimulus.

Store-and-forward audio and video recording. In the online-led and FTF-led assessments, the online SLP was able to prompt the system to record audio clips of the child's responses via the headset microphone. The audio data were captured and compressed at 70 Kbs using the Microsoft Windows Media Encoder Codec (Version 9).

During the FTF-led assessment, it was necessary for the online SLP to record video of the child pointing to items in the stimulus book during the Concepts and Following Directions subtest. The online SLP prompted the system to record these responses via one participant Web camera and the headset microphone. High-quality data was captured at 640 × 480 pixels resolution at 25 frames per second, compressed using the Microsoft Windows Media Encoder Codec (Version 9), with a bit rate of 235 Kbs for video and 70 Kbs for audio. This feature enabled the online clinician to view high-quality footage of the participant that is not possible via the 128-Kbs videoconferencing, thus facilitating more reliable scoring of the assessments. Immediately following the audio or video recording of each response, the system software forwarded the file to the online SLP's computer and stored the file on the participant's computer. The online SLP could then play back the forwarded file immediately or following the session.

Store-and-forward touch data recording. The participant's computer monitor used in the study was an analogue capacitive touch screen that enabled recording of the participant's responses to stimuli displayed on the screen. The touch screen was used in the online-led assessment of the Concepts and Following Directions subtest only, where the child was required to point to items on the screen. Following the child's response, the touch screen data were recorded and a text file of the coordinates of the touch was stored on the participant's computer and forwarded to the clinician's computer. The touch data sent from the participant's computer were loaded onto the reciprocal image on the SLP's computer for immediate analysis or could be reloaded following the assessment session. The participant was able to practice using the touch screen during the demonstration and trial items of the subtest. The touch screen was unable to record two simultaneous touch responses. Therefore, the child was encouraged to use one finger to point to the screen unless he or she thought the question required the use of two fingers. For questions requiring the child to point to two items at the same time, the FTF SLP recorded the response and reported the response to the online SLP following the assessment. The FTF SLP reported only how the participant responded and not whether the response was correct or incorrect. In a real-world setting, it is likely that there would always be an adult present with the child who could clarify such ambiguous responses for an online SLP.

Data security. Functions were built into the system to maximize the security of the data transferred across the Internet (Hill et al., 2006). These were developed according to NetMeeting security guidelines (Hayes & Pitsenbarger, 2003).

Statistical Analysis

Online and FTF rating comparison. Before statistical analysis, statistical procedures were used to determine whether the data for the total raw scores and scaled scores were normally distributed. Normally distributed data were found for all parameters based on the results of the Kolmogorov-Smirnov test of normality ($p > .01$ for each parameter) and nonsignificant skewness and kurtosis at $p < .01$ significance. As the scores were normally distributed and based on interval scales, parametric statistics were performed.

In order to determine whether the administration environment (online-led, FTF-led) as selected by the randomization code had an impact on the accuracy of the assessments, a statistical test of difference was performed. This statistic, an independent *t* test, was calculated on the difference scores between the online rating and the FTF rating obtained for each administration environment. If no difference was found, it would be assumed that the administering environment did not influence the accuracy of the online assessment and data would be pooled across administration environments. The significance level for this test was established by applying a Bonferroni procedure to correct for multiplicity of tests (Type 1 error). As there were eight measures, the alpha level was set at .006 (approximately .05/8). The differences for the two administration environments were not found to be significantly different ($p > .006$; see Table 1) on any subtest, and the 95% confidence intervals indicate negligible differences between the means. Therefore, the online and FTF scores obtained from the two administration environments were collapsed in the subsequent analyses.

The results obtained on the various CELF–4 measures by the online and FTF SLPs were compared using several statistical processes. The total raw scores and scaled scores on each of the four CELF–4 subtests for the online and FTF scoring conditions were compared using paired *t* tests together with 95% confidence intervals. The alpha level was set at .006 (Bonferroni correction). Cohen's unweighted kappa (Cohen, 1960) was used to calculate agreement beyond the expected level of chance between the online and FTF ratings for the comparison of each individual test item up to each participant's discontinue point on the Concepts and Following Directions and Word Structure subtests. The unweighted kappa statistic was used because the response decision was dichotomous. Items that were not scored by both SLPs were excluded from analysis. The criteria for level of agreement were based on Altman (1991), whereby κ values < .2 at the 95% confidence interval indicate poor strength of agreement, .21–.40 = fair, .41–.60 = moderate, .61–.80 = good, and .81–1.00 = very good agreement. The weighted kappa statistic (quadratic weighting; Fleiss & Cohen, 1973) was used to determine the level of agreement between the online and FTF ratings on each individual item up to the discontinue points on the Recalling Sentences and Formulated Sentences subtests, as these ratings consisted of more than two categories. Unlike the unweighted kappa, the weighted kappa takes into account the magnitude of the disagreement, thus giving partial credit for close agreements (Maclure & Willett, 1987). The Altman (1991) criteria were also used for the weighted kappa analysis. The weighted kappa statistic (quadratic weighting) was calculated to determine agreement between the online and FTF subtest total raw scores, scaled scores, core language score, and severity level.

Intrarater and interrater reliability. Intrarater reliability was determined for a sample of online ratings. Eight online-led assessments were selected via a random numbers table, and these assessments were rerated by the online SLP. To control for carryover effects, the second rating was completed at least 4 weeks after the initial rating (Yoon, Starr, Perkins, Bloom, & Sie, 2006). The recordings were viewed in the same manner as for the original ratings. Interrater reliability was determined for a random sample of eight online-led assessments, whereby an independent rater scored the online recordings. Intra- and interrater agreement were calculated using intraclass correlation coefficients ($ICC_{2,1}$; Shrout & Fleiss, 1979) on the individual item scores, subtest raw scores, scaled scores, core language score, and severity level. The criteria for this statistic was set at ICC <.4 = poor agreement, .4–.6 = moderate, .6–.8 = good, and >.8 = very good agreement (Fleiss, 1981).

Intra- and interrater reliability were also determined for the FTF scoring of a sample of eight FTF-led assessments, selected by

Table 1. Comparison of difference scores across the online-led and the face-to-face (FTF)-led assessment.

Subtest	*Difference score in online-led condition*[a]		*Difference score in FTF-led condition*[b]		*95% CI of the difference*			
	M	SD	M	SD	*Lower*	*Upper*	t	p*
Concepts and Following Directions								
Total raw score	–0.15	0.38	–0.33	0.89	–0.38	0.74	0.67	.51
Scaled score	–0.08	0.28	–0.08	0.29	–0.23	0.24	0.06	.96
Word Structure								
Total raw score	–0.15	0.99	0.00	1.48	–1.19	0.88	–0.31	.76
Scaled score	–0.08	0.76	0.00	0.60	–0.65	0.49	–0.28	.78
Recalling Sentences								
Total raw score	–0.46	1.39	–0.08	2.02	–1.80	1.05	–0.55	.59
Scaled score	–0.08	0.28	–0.17	0.58	–0.30	0.48	0.49	.63
Formulated Sentences								
Total raw score	0.08	1.55	0.58	2.15	–2.05	1.04	–0.68	.50
Scaled score	0.08	0.76	0.25	1.06	–0.93	0.58	–0.47	.64

Note. The difference score was calculated using the following formula: online rating – FTF rating. CI = confidence interval.

[a]n = 13, [b]n = 12.

*nonsignificant difference for all comparisons, $p > .006$.

random numbers table. The FTF clinician rerated the assessments at least 4 weeks following the initial rating, and the independent rater scored the assessments in the same manner as the original rating. According to the criteria for ICC, the intrarater reliability was very good for the FTF ratings on all measures (individual items, ICC = .87–.99; total raw scores = .99–1.00; scaled scores = .97–1.00; core language score = >.99; severity level = 1.00). Very good interrater reliability was determined for all FTF scores except for the formulated sentences individual item score, for which good agreement was found (individual items ICC = .74–.97; total raw scores = .95–>.99; scaled scores = .94–.99; core language score = .98; severity level = .97). The interrater reliability reported in the manual of the Australian standardization of the CELF–4 was reported to be ICC = .99–1.00 for the subtest total raw scores (Semel, Wiig, & Secord, 2006). As the clinicians performed similarly to those in the reliability data of the CELF–4, the SLPs have not been included as a separate factor in the online and FTF scoring comparison.

Informal analysis of technical feasibility and child-related issues. During each assessment, the principal researcher recorded observations about any technical difficulties or child-related problems encountered during the session.

RESULTS

Comparison of Online and FTF Ratings

Subtest scores. The paired t tests revealed no significant difference between the online and FTF total raw scores and scaled scores ($p > .006$) for all subtests. The results of this analysis and the 95% confidence intervals for the raw scores and scaled scores are provided in Table 2. Very good agreement was determined between the two raters for the total raw scores and scaled scores (κ > .90) for all subtests (Table 3). The difference between the online and FTF scaled scores was no greater than 1 point for the Concepts and Following Directions and Recalling Sentences subtests, and no greater than 2 points for the Word Structure and Formulated Sentences subtests, for any participant.

Individual item scores. The kappa and weighted kappa analyses revealed very good agreement between the online and FTF ratings for the individual item scores on all subtests (κ = .88–.98, see Table 3 for each subtest's results). It should be noted that some individual items were excluded from the analysis. For the Concepts and Following Directions subtest, 216 items were not scored by either rater as they were past the subtest discontinue points. One item was not video-recorded by the online SLP during an FTF-led assessment, and therefore could not be scored. There was no discontinuation point for Word Structure. Two individual Word Structure items were excluded from the analysis as they were not audio-recorded by the online SLP during an FTF-led assessment (due to human error), and therefore were not able to be scored. For Recalling Sentences, there were 312 individual items that were not included in the analysis. Of these, 305 were not scored by either rater as they were past the subtest discontinue points. However, six individual items were only rated by the online SLP, as the FTF rater had already reached the discontinue point. One item was only scored by the FTF SLP because the online SLP had reached the discontinue point. These discrepancies occurred on three separate FTF-led assessments. Finally, for Formulated Sentences, 160 individual items on the subtests were not included in the weighted kappa analysis. Of these, 156 items were not scored by either the online or FTF SLP because they were past the discontinue point. Two items from one online-led assessment were only scored by the FTF rater due to a discrepancy in the discontinue points. Two items were not scored by the online rater as recordings were not made and the items were not rated during the assessment. The

Table 2. Comparison of online and FTF scores on the subtests of the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003).

Subtest	*Online*[a] M	SD	*FTF*[a] M	SD	*95% CI of the difference* *Lower*	*Upper*	t	p*
Concepts and Following Directions								
Total raw score	21.28	12.22	21.52	12.46	−0.51	0.03	−1.81	.08
Scaled score	5.56	2.92	5.64	2.98	−0.19	0.03	−1.45	.16
Word Structure								
Total raw score	17.00	8.72	17.08	8.80	−0.58	0.42	−0.33	.75
Scaled score	5.96	3.77	6.00	3.82	−0.32	0.24	−0.30	.77
Recalling Sentences								
Total raw score	25.72	16.73	26.00	16.61	−0.98	0.42	−0.83	.42
Scaled score	5.04	3.25	5.16	3.38	−0.30	0.06	−1.37	.19
Formulated Sentences								
Total raw score	18.24	12.97	17.92	12.78	−0.44	1.08	0.87	.39
Scaled score	6.68	3.95	6.52	4.20	−0.21	0.53	0.89	.38

[a]$n = 25$.

*nonsignificant difference for all comparisons, $p > .006$.

missing recordings occurred in one online-led and one FTF-led assessment.

Core language score and severity level. Very good levels of agreement were also found for the core language score (κ = .99) and the severity level (κ = .99). The difference between the online and FTF core language score was no greater than 5 points for any participant. There was complete agreement on the severity level for 24 out of the 25 (96%) participants, with the remaining participant only differing by one severity level.

Intra- and interrater reliability. According to the criteria for ICC, the intrarater reliability was very good (>.80) for the online ratings on all measures (see Table 4). ICCs for the individual item scores ranged from .91 to .99. ICCs of at least .97 were achieved for all subtest total raw scores and scaled scores. An ICC of .99 was

Table 3. Results of kappa or weighted kappa analysis for online and FTF ratings on all CELF–4 subtests.

Subtest	n	κ
Concepts and Following Directions		
Individual items	1133	.98[a]
Total raw score	25	>.99[a]
Scaled score	25	.99[a]
Word Structure		
Individual items	798	.93[a]
Total raw score	25	.98[a]
Scaled score	25	.99[a]
Recalling Sentences		
Individual items	488	.96[a]
Total raw score	25	.99[a]
Scaled score	25	.99[a]
Formulated Sentences		
Individual items	440	.88[a]
Total raw score	25	.99[a]
Scaled score	25	.97[a]
Core language score	25	.99[a]
Severity level	25	.99[a]

Note. Statistics for individual items were calculated on different *n*s due to test discontinue rules and missing data.

[a]Very good agreement (κ > .81).

Table 4. Intra- and interrater reliability for online CELF–4 assessments.

Subtest	*Intrarater reliability* n	*ICC*	*Interrater reliability* n	*ICC*
Concepts and Following Directions				
Individual items	361	.99	343	.98
Total raw score	8	>.99	8	>.99
Scaled score	8	1.00	8	1.00
Word Structure				
Individual items	253	.93	251	.89
Total raw score	8	.98	8	.96
Scaled score	8	.98	8	.92
Recalling Sentences				
Individual items	154	.98	146	.93
Total raw score	8	>.99	8	>.99
Scaled score	8	.99	8	.99
Formulated Sentences				
Individual items	128	.91	127	.84
Total raw score	8	.99	8	.99
Scaled score	8	.97	8	.98
Core language score	8	.99	8	.98
Severity level	8	1.00	8	1.00

Note. Statistics for individual items were calculated on different *n*s due to test discontinue rules and missing data. ICC = intraclass correlation coefficient.

achieved for the intrarater reliability on the core language score, and complete agreement (ICC = 1.00) was obtained for the severity level. Very good interrater reliability was found for the online individual item scores for each subtest (ICC range = .84–.98; refer to Table 4). Interrater reliability was similarly very good for the total raw scores (ICC = .96–>.99), scaled scores (ICC = .92–1.00), core language score (ICC = .98), and severity level (ICC = 1.00).

Informal analysis of technical and child-related issues. A number of logistical and technical issues and difficulties caused by child-related factors were encountered during both the online-led and FTF-led assessments. These issues are summarized in Table 5. The majority of difficulties were observed in only a small number of assessments. It should be noted that some issues observed in the FTF-led assessments were an artifact of the study environment and would not normally occur in an FTF assessment.

DISCUSSION

This study aimed to determine the validity and reliability of online assessment of pediatric language using the core subtests of a standardized tool, the CELF–4. The results of the present investigation have demonstrated the concurrent validity of the online scoring of the core language subtests of the assessment (Kazdin, 2003). The online ratings of the four core language subtests for children ages 5–8 did not differ significantly from the FTF ratings, and the confidence intervals calculated revealed that the range of possible differences in the online and FTF scores were negligible. The validity of the online assessment was also demonstrated by the very good strength of agreement between the online and FTF individual item scores, total subtest scores and scaled scores, core language scores, and severity levels. Further, the online assessment was found to have both high intra- and interrater reliability. Although the findings of the present study cannot be generalized beyond the four core language subtests and the SLPs who participated in this study, it is likely that future research may show similar results for other subtests and raters.

Before the present research, only two studies had attempted to deliver a formal language assessment via telehealth (Cole et al., 1986; Fairweather et al., 2004). The study most similar to the current investigation was that conducted by Fairweather and colleagues (2004), which included the assessment of 13 school-age children (ages 6–14) on the CELF–3 via videoconferencing. In this study, each participant was simultaneously rated on the CELF–3 by a remote and an FTF SLP, and the ratings of the two SLPs were compared. However, unlike the present study, all assessments were conducted only by the remote SLP. The FTF SLP manipulated test

Table 5. Observations of technical and child-related issues during the assessments.

Issue	*Description*	*Impact on assessments*
Technical issues		
Equipment size	The headset was too big and fell off during assessment.	Occurred in one online-led assessment. It was distracting for the child and may have impacted performance.
Bandwidth/network congestion	The combination of low bandwidth and high network congestion caused distortion in the audio signal and/or audio breakup.	Impacted the child's ability to hear some general directions and parts of the Concepts and Following Directions subtest in three online-led assessments.
Touch screen	Multiple touches could not be recorded.	Impacted two items on the Concepts and Following Directions subtest for all 13 online-led assessments. The FTF speech-language pathologist was able to clarify what had been pointed to when requested.
	No response was recorded when the child did not the touch screen.	Affected one item in one assessment. The online speech-language pathologist did not realize the child had responded.
Lighting	Lighting caused overexposure in the online video recordings of the test book.	Impacted two FTF-led assessments. Made it difficult for the online speech-language pathologist to distinguish certain pictures only.
Positioning of participant	The child was positioned in front of the stimulus book so it was difficult for the online SLP to view responses in the Concepts and Following Directions subtest.	Occurred in one FTF-led assessment.
Child-related issues		
Intelligibility	The participant had low speech volume and intelligibility.	Occurred in two online-led and one FTF-led assessments. Made it difficult for both online and FTF speech-language pathologists to score the expressive language tasks from recordings.
Interaction with online speech-language pathologist	The participant did not request clarification or repetition when there was breakup in audio during the videoconference.	Occurred in one online-led assessment during the Concepts and Following Directions subtest. The FTF speech-language pathologist was able to provide assistance.
	The participant did not request to self-correct.	Occurred in one online-led assessment during the Concepts and Following Directions subtest only.
Attention and concentration	The child was tired and restless after school.	Occurred in one FTF-led assessment.
Motor skills	The child had difficulty operating the touch screen; did not touch items on the screen directly and did not lift his/her finger from the screen.	Occurred with one participant during the online-led assessment. Extra prompting was required. This issue did not impact the speech-language pathologists' ability to score the assessment.

materials but otherwise scored the assessments passively. Fairweather et al. reported "reasonable levels of reliability," with complete agreement for the individual subtest scores for 69% of the participants. However, it was not stated whether the subtest raw scores or scaled scores were compared. These results suggested that exact agreement between raters is difficult to achieve.

Although the current study found very good agreement between the two assessors for the total raw scores and scaled scores on each subtest, it was also necessary to determine the level of agreement for the individual item scores. If one purpose of the assessment is to determine which particular language features the child can produce or understand, then the ability to score each item correctly is important. A very high level of agreement between raters would be expected for the Concepts and Following Directions and Recalling Sentences subtests, as these are scored objectively, whereas the Word Structure and Formulated Sentences subtests require subjective judgment from the examiners (Semel et al., 2003). For the individual items on the subtests, the highest level of agreement ($\kappa = .98$) was determined for the Concepts and Following Directions subtest. However, this result indicates that complete agreement was not achieved. Discrepancies between the online and FTF ratings of the Concepts and Following Directions subtest occurred mainly during FTF-led assessments. The discrepancies were due to issues such as problems with the positioning of the participant in relation to the stimulus book in one assessment, and with lighting and exposure of the Web cameras in two assessments. These issues made it difficult for the online SLP to see the participants' responses. The online-led assessment procedure used the touch screen to record the responses on the Concepts and Following Directions subtest, therefore overcoming any limitations of the Web cameras and enabling more accurate scoring in the online environment. This finding suggests that, in future studies of this type, a touch screen may be better for recording responses on receptive language subtests than having a camera directed at the stimulus book.

Slightly lower agreement was found for the individual items on the Recalling Sentences subtest ($\kappa = .96$) than for the Concepts and Following Directions subtest. It was felt that differences in the recording of responses were largely due to some participants' level of intelligibility and volume rather than problems with the online environment per se. Poor intelligibility made it difficult for both the online and FTF SLPs to record participant responses. Some participants spoke with a low volume and/or had a co-occurring speech delay, which reduced their intelligibility. Interrater differences due to difficulties discerning participant's speech were also noted by Fairweather and colleagues (2004) in their investigation of the assessment of children on the CELF–3 via videoconferencing. Therefore, less than perfect agreement due to difficulty hearing the child over the telehealth system is not unusual in telehealth studies, suggesting that factors such as the age and intelligibility of the child need to be taken into consideration when conducting online assessments. It is possible that video-recording instead of audio-recording the child's responses may have improved the online SLP's ability to accurately transcribe his or her responses and score Recalling Sentences and the other expressive language subtests.

The Word Structure and Formulated Sentences subtests required the SLPs to make a subjective judgment about the correctness of the child's response. The levels of agreement obtained on the individual items for Word Structure and Formulated Sentences were slightly lower than for Concepts and Following Directions and Recalling Sentences, being .93 and .88, respectively. Although the SLPs consistently followed the scoring guidelines set out in the test manual, differences in judgment remained. Interrater difference in scoring Formulated Sentences due to subjective differences was also reported by Fairweather and colleagues (2004). As with Recalling Sentences, the ability of the online rater to accurately transcribe and score the Word Structure and Formulated Sentences subtests depended on the quality of the audio signal of the recording. Reduced speech volume and intelligibility affected both the online and FTF raters' ability to hear the participants' responses. Difficulty discerning the child's responses may be a concern for Word Structure if information regarding the use of particular morphological structures is required. In the current study, a closer examination revealed more cases in which the online and FTF clinicians disagreed in their scoring of regular plural (*books*), third person singular (*reads*), irregular plural (*children*), and auxiliary + *ing* (*they are swinging*), than for the other items on the Word Structure subtest. In a clinical situation, further informal assessment may be needed if there is doubt about the accuracy of a particular structure. However, despite differences between the scoring of some individual items, the overall agreement levels for the individual items were very good.

It was also necessary to determine whether any differences between the online and FTF ratings had an impact on the core language score and severity level. Although a complete CELF–4 assessment as well as other assessment and sampling procedures would be necessary for a comprehensive diagnosis of a language impairment, the core language score can be used to determine whether a language disorder is present and to measure the child's overall language ability (Semel et al., 2003). In the present study, there was very good agreement in the core language scores determined by the online and FTF SLPs ($\kappa = .99$). The online and FTF ratings of severity levels, as determined by the core language score, were found to be in complete agreement for all but one participant. For this participant, the severity levels identified by the assessors were mild and moderate. However, the core language scores calculated by each SLP for this participant were 76 and 79, respectively, and were, therefore, relatively close. The results of the present investigation indicated that assessment on the CELF–4 in an online scoring environment can identify the presence of a language disorder and determine a severity level based on the core language score.

Intrarater reliability for the online scoring was found to be very good for all parameters tested; with ICCs (.91–1.00) being very similar to those originally calculated for these clinicians in the FTF scoring environment (.87–1.00). This finding suggests that for the SLPs in this investigation, assessment in the online environment was as reliable as the FTF assessment. The online assessment also demonstrated very good interrater reliability across all measures (ICC ranging from .84 to 1.00). The Australian standardization version of the CELF–4 reported on interscorer reliability between two ratings of 69 assessments (Semel et al., 2006). Intraclass correlation was calculated for the total subtest scores and was reported to be either .99 or 1.0 for all subtests. The ICCs for the subtest raw scores in the current study were similar to those reported in the manual (.96–>.99). This finding suggests that in the online condition, the raters in the current study were as reliable as those used in the standardization of the CELF–4.

In the present study, it was interesting to examine the impact the child's age had on the effectiveness of the telehealth assessment. The online-led assessment protocol appeared to be effective in assessing children across the entire 5- to 9-year age range. The youngest child to be allocated to the online-led assessment group was 5;6. Fairweather and colleagues (2004) commented that speech-language pathology services via videoconferencing would have the most potential for children older than 6 years of age, as younger children may need furniture modifications, may have difficulty maintaining focus during the assessment, or may have difficulty interacting with an SLP over the videoconference. They recommended that further research with children ages 0–6 be conducted (Fairweather et al., 2004). In the present study, the investigators chose to administer only four subtests so that the assessment protocol would not be too long. The children were generally able to maintain focus throughout the assessment as small breaks were allowed. The researchers felt that a longer assessment protocol might have been difficult for a younger child. However, this would also be the case for any traditional FTF assessment in this population. A study of the telepsychiatry assessment in 23 children ages 4–16 via a PC-based videoconferencing system found that children had a tendency to move about the room and were distracted by equipment wires and buttons (Elford et al., 2000). In the present study, no such difficulty was encountered, possibly due to the presence of the FTF SLP or interest in the assessment software. Similarly, the children included in the present study did not have any identified psychological conditions and therefore may have presented with better behavior and attention than those included in Elford and colleagues' (2000) investigation. Further investigation of the ability to assess children with attention deficits or behavioral problems via telehealth is necessary.

Human factors are important to consider when evaluating a telehealth application. Unlike Fairweather et al. (2004), the researchers of the current study did not find rapport building via the telehealth system difficult, and were satisfied overall with the interaction. No formal clinician satisfaction questionnaire was conducted in this study, but one could be included in future investigations. Although Eriks-Brophy, Quittenbaum, Anderson, and Nelson (2008) found that the younger children assessed via videoconferencing appeared shyer than the older children, they felt that the technology was not the sole reason behind this behavior, as these children behaved similarly in an FTF situation. In the present investigation, some children did not request clarification or repetition when there was breakup of the audio during the videoconference, or request permission to correct themselves. It is possible that this behavior may have also occurred in an FTF assessment. It is also possible, however, that in an FTF situation, the SLP may have been better able to observe the child's nonverbal cues that he or she did not understand or wanted to self-correct. In light of these observations, it would be beneficial to carry out further investigation as to whether a child interacts and performs in the same way in an online assessment as in a traditional FTF assessment.

For younger or smaller participants, the investigators agree with Fairweather and colleagues (2004) regarding the need for equipment modification. The headset microphone was too large for some of the children, although this seemed to be distracting for only one child. The seat had to be raised so that the children could use the touch screen. In light of the observations that one child with delayed motor skills had difficulty operating the touch screen, it is recommended that the impact that children's motor skills have on the ability to conduct the online assessment is investigated further. Alternative response formats could be incorporated into an Internet-based telehealth system for children with physical limitations. For example, scanning and trackball responses to items presented on a computer for the PPVT–R have been found to produce no difference in performance than traditional administration of the test (Haaf et al., 1999). Other equipment modifications, such as child-sized headphones or an echo-cancelling microphone and speaker system, should also be considered.

The present investigation sought to determine the feasibility of the telehealth technology in assessing young children's language. There were a number of benefits of the technology used in this study. One major benefit of using a PC-based telehealth system over the systems used in previous studies of standardized pediatric language assessment via telehealth (Cole et al., 1986; Fairweather et al., 2004) was that the online-led assessments were completely administered by the remote SLP. In previous studies, the FTF SLP was required to manipulate the test materials. The system used in the current study allowed the materials to be presented in computerized format, and therefore there was no need to transport the materials to the remote location or to train an adult at the remote location to manipulate them. The capacity of the PC-based system to make store-and-forward recordings of the child's responses was also an advantage over other systems. It provided recordings for later analysis as well as a permanent record of the child's performance. Immediate access to the audio and video recordings is an advantage as it may be necessary for the online SLP to check the child's productions during the assessment, particularly during periods of audio or video breakup.

There were also a number of technological difficulties observed during the assessments; however, many of these occurred in only a minority of assessments. The videoconferencing system used in the present investigation has demonstrated that even while running at a low bandwidth (128 Kbs), an acceptable level of agreement between online and FTF ratings is possible. However, the low bandwidth as well as a high amount of network congestion limited the capacity to provide optimal audio and visual quality during the videoconference. Video recordings of the stimuli for the Recalling Sentences subtest were included to provide a consistent stimulus that would not be affected by audio breakup. However, in order to facilitate administration of these parts of the assessment, no such recordings were made for the other subtests. Some audio breakup did occur during three assessments, particularly during Concepts and Following Directions, which may be viewed as a concern. In the present investigation, recordings were taken of the participant's responses to overcome limitations of the poor audio quality of the videoconference. However, this meant that the online SLP had to play back the recordings following the assessment, which increased the time it took to score the assessment. These limitations of the low bandwidth were found in other applications using a similar system (Hill et al., 2008; Theodoros et al., 2008). Increasing the bandwidth may alleviate the audio breakup problem and improve the visual quality of the videoconference (Hill et al., 2009). The presentation of recordings of all auditory stimuli would also overcome the issues of audio breakup during the assessment.

Another technological limitation was the inability of the touch screen to record multiple simultaneous responses for the Concepts and Following Directions subtest. An assistant or parent at the student site could be trained to record the responses when a child points to more than one place at the same time. However, parents may record a correct response when the child is actually incorrect (Wilson et al., 2006). Advancing technology has meant that multipoint touch screens are now being developed, allowing the tracking of more than one finger on the screen. A multipoint touch screen would most accurately record the responses to a subtest such as Concepts and Following Directions. Despite some limitations in the technology used in this investigation, overall, the technology was adequate for the assessment of children's language on the core subtests of the CELF–4.

Conclusion

The current study revealed no significant difference in the scores on the CELF–4 core language subtests obtained by an online and an FTF SLP. Furthermore, very good agreement was found between the two ratings on all measures. High intra- and interrater reliability for the online assessments support the feasibility of using this Internet-based telehealth system for assessing children's language. Although the technology was adequate for this assessment, some modifications, such as increasing bandwidth and including a multipoint touch screen, would improve the system's effectiveness. Further research should aim to validate the entire CELF–4 and include larger sample sizes in order to explore the effect of the child's age and severity on the online assessment, and to look at factors such as behavior, attention, and physical ability. Future research should also focus on the way the child interacts and responds to assessment in the online environment. Once these factors are investigated, the capacity to provide this telehealth assessment in schools, clinics, or homes should also be explored. The present research supports the use of telehealth in the assessment of children's language.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Roy Anderson for his assistance in software development; Jasmin Cowles, Chiara Wall, and Eva Loh for data collection; and Asad Khan for statistical advice. The authors also acknowledge the University of Queensland Health and Rehabilitation Clinics and the schools involved for assistance in participant recruitment. Thank you to all the children and their families for their time and enthusiasm in taking part in the study.

REFERENCES

Altman, D. G. (1991). *Practical statistics for medical research*. London, England: Chapman & Hall.

American Speech-Language-Hearing Association. (1985, November). 1985 National Colloquium on Underserved Populations report. *Asha, 27*(11), 31–35.

American Speech-Language-Hearing Association. (2005). *Speech-language pathologists providing clinical services via telepractice: Position statement* [Position Statement]. Available from www.asha.org/policy.

American Speech-Language-Hearing Association. (2007). *ASHA SLP Health Care Survey 2007: Workforce/work conditions*. Rockville, MD: Author.

American Speech-Language-Hearing Association. (2008). *Schools Survey report: Workforce and work conditions trends 2000–2008*. Rockville, MD: Author.

Australian Bureau of Statistics. (2003). *Australian social trends 2003* (Catalogue no. 4102.0). Canberra, Australia: Author. Retrieved from http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/70AC02A724F4033FCA256D3A0006B054/$File/41020_2003.pdf.

Baker, L., & Cantwell, D. P. (1987). A prospective psychiatric follow-up of children with speech/language disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 546–553.

Beitchman, J. H., Wilson, B., Brownie, E. B., Walters, H., Inglis, A., & Lancee, W. (1996). Long-term consistency in speech/language profiles: II. Behavioural, emotional, and social outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 815–825.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 31,* 1027–1050.

Blum-Harasty, J. A., & Rosenthal, J. B. M. (1992). The prevalence of communication disorders in children: A summary and critical review. *Australian Journal of Human Communication Disorders, 20,* 63–80.

Brennan, D. M., Georgeadis, A. C., Baron, C. R., & Barker, L. M. (2004). The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemedicine Journal and e-Health, 10,* 147–154.

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language & Communication Disorders, 39,* 303–324.

Cantwell, D. P., & Baker, L. (1987). Clinical significance of childhood communication disorders: Perspectives from a longitudinal study. *Journal of Child Neurology, 2,* 257–264.

Chezik, K. H., Pratt, J. E., Stewart, J. L., & Deal, V. R. (1989, January). Addressing service delivery in remote/rural areas. *Asha, 31*(1), 52–55.

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders: A follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46,* 128–149.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cole, C., Martin, V., Moody, J., & Miller, B. (1986). Speech pathology at a distance. *Australian Communication Quarterly, 2,* 17–19.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: AGS.

Elford, R., White, H., Bowering, R., Ghandi, M., Maddigan, B., St John, K., . . . Battock, A. (2000). A randomized, controlled trial of child psychiatric assessments conducted using videoconferencing. *Journal of Telemedicine and Telecare, 6,* 73–82.

Eriks-Brophy, A., Quittenbaum, J., Anderson, D., & Nelson, T. (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clinical Linguistics and Phonetics, 2,* 589–609.

Fairweather, C., Parkin, M., & Rozsa, M. (2004). *Speech and language assessment in school-aged children via videoconferencing*. In B. E. Murdoch, J. Goozee, B. Whelan, & K. Docking (Eds.), Proceedings of the 26th World Congress of the International Association of Logopaedics and Phoniatrics (IALP) [CD-ROM]. Melbourne, Australia: Speech Pathology Australia.

Fisher, H. B., & Logemann, J. A. (1971). *The Fisher-Logemann Test of Articulation Competence*. Austin, TX: Pro-Ed.

Fitzgerald, K., Hornsby, D., & Hudson, L. (2000). *A study of allied health professionals in rural and remote Australia*. Bathurst, New South Wales, Australia: Services for Australian Rural and Remote Allied Health.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33,* 613–619.

Forducey, P. G. (2006, August 15). Speech telepractice program expands options for rural Oklahoma schools. *The ASHA Leader, 11*(10), 12–13.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation* (2nd ed.). Circle Pines, MN: AGS.

Haaf, R., Duncan, B., Skarakis-Doyle, E., Carew, M., & Kapitan, P. (1999). Computer-based language assessment software: The effects of presentation and response format. *Language, Speech, and Hearing Services in Schools, 30,* 68–74.

Hayes, J. M., & Pitsenbarger, T. H. (2003). *Microsoft NetMeeting 3.0 security assessment and configuration guide*. Retrieved from http://nsa2.www.conxion.com/support/guides/sd-6.pdf.

Hill, A. J., & Theodoros, D. (2002). Research into telehealth applications in speech-language pathology. *Journal of Telemedicine and Telecare, 8,* 187–196.

Hill, A. J., Theodoros, D., Russell, T., & Ward, E. (2008). Using telerehabilitation to assess apraxia of speech in adults. *International Journal of Language & Communication Disorders,* Advance online publication. doi:10.1080/13682820802350537.

Hill, A. J., Theodoros, D. G., Russell, T. G., Cahill, L. M., Ward, E. C., & Clark, K. M. (2006). An Internet-based telerehabilitation system for the assessment of motor speech disorders: A pilot study. *American Journal of Speech-Language Pathology, 15,* 45–56.

Hill, A. J., Theodoros, D. G., Russell, T. G., Ward, E. C., & Wootton, R. (2009). The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology, 23,* 627–642.

Humphreys, J., Hegney, D., Lipscombe, J., Gregory, G., & Chater, B. (2002). Whither rural health? Reviewing a decade of progress in rural health. *Australian Journal of Rural Health, 10,* 2–14.

Iacano, T., Johnson, H., Humphreys, J., & McAllister, L. (2007). Recruitment of speech pathologists into positions considered less attractive. *Advances in Speech-Language Pathology, 9,* 204–212.

Jessup, B., Ward, E., Cahill, L., & Keating, D. (2008). Prevalence of speech and/or language impairment in preparatory students in northern Tasmania. *International Journal of Speech-Language Pathology, 10,* 364–377.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., . . . Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–760.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.

Lambier, J., & Atherton, M. (2003). *General membership survey: Speech Pathology Association of Australia 2003*. Retrieved from http://www.speechpathologyaustralia.org.au/Content_Secure.aspx?p=165.

Lemaire, E. D., Boudrias, Y., & Greene, G. (2001). Low-bandwidth, Internet-based videoconferencing for physical rehabilitation consultations. *Journal of Telemedicine and Telecare, 7,* 82–89.

Maclure, M., & Willett, W. C. (1987). Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology, 126,* 161–169.

National Rural Health Alliance. (2004). *Under pressure and undervalued: Allied health professionals in rural and remote areas*. Retrieved from http://nrha.ruralhealth.org.au/?IntCatId=14.

O'Callaghan, A. M., McAllister, L., & Wilson, L. (2005). Barriers to accessing rural paediatric speech pathology services: Health care consumers' perspectives. *Australian Journal of Rural Health, 13,* 162–171.

Pickering, M., McAllister, L., Hagler, P., Whitehill, T. L., Penn, C., Robertson, S. J., & McCready, V. (1998). External factors influencing the profession in six societies. *American Journal of Speech-Language Pathology, 7*(4), 5–17.

Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. (2009). *World urbanization prospects: The 2007 revision population database*. Retrieved from http://esa.un.org/unup.

Renfrew, C. E. (1971). *Action Picture Test*. Oxford, England: North Place, Headington.

Savard, L., Borstad, A., Tkachuck, J., Lauderdale, D., & Conroy, B. (2003). Telerehabilitation consultations for clients with neurologic diagnoses: Cases from rural Minnesota and American Samoa. *NeuroRehabilitation, 18,* 93–102.

Scheideman-Miller, C., Clark, P. G., Smeltzer, S. S., Carpenter, J., Hodge, B., & Prouty, D. (2002). Two year results of a pilot study delivering speech therapy to students in a rural Oklahoma school via telemedicine. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (Volume 6, p. 161b). Los Alamos, CA: IEEE Computer Society Press.

Semel, E. M., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Semel, E. M., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.

Semel, E. M., Wiig, E. H., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals* (4th-Australian Standardised Edition). Marrickville, New South Wales, Australia: Harcourt Assessment.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. C., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407–418.

Theodoros, D., Hill, A., Russell, T., Ward, E., & Wootton, R. (2008). Assessing acquired language disorders in adults via the Internet. *Telemedicine and e-Health, 14,* 552–559.

Victorian Department of Human Services. (2004). *Children's speech and language services: Eastern metropolitan region*. Melbourne, Victoria, Australia: Eastern Metropolitan Regional Office.

Waite, M. C., Cahill, L. M., Theodoros, D. G., Busuttin, S., & Russell, T. G. (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare, 12*(Suppl. 3), 92–94.

Ward, L., White, J., Russell, T., Theodoros, D., Kuhl, M., Nelson, K., & Peters, I. (2007). Assessment of communication and swallowing function post laryngectomy: A telerehabilitation trial. *Journal of Telemedicine and Telecare, 13*(Suppl. 3), 88–91.

Wilson, L., Lincoln, M., & Onslow, M. (2002). Availability, access, and quality of care: Inequities in rural speech pathology services for children and a model for redress. *Advances in Speech-Language Pathology, 4,* 9–22.

Wilson, L., McAllister, L., Atkinson, J., & Sefton, P. (2006). Internet based telehealth in speech pathology: Results of preliminary trials. In F. Pinciroli (Ed.), *Proceedings of the Second IASTED International Conference on Telehealth* (pp. 131–136). Anaheim, CA: ACTA.

Yoon, P. J., Starr, J. R., Perkins, J. A., Bloom, D., & Sie, K. C. Y. (2006). Interrater and intrarater reliability in the evaluation of velopharyngeal insufficiency within a single institution. *Archives of Otolaryngology—Head and Neck Surgery, 132,* 947–951.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry, 43,* 635–645.

Received December 23, 2008
Revision received July 13, 2009
Accepted December 14, 2009
DOI: 10.1044/0161-1461(2009/08-0131)

Contact author: Monique Waite, The University of Queensland, Division of Speech Pathology, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: m.waite@uq.edu.au.