To the faculty of Washington State University:

The members of the Committee appointed to examine the thesis of ALYSSA ANDERSON find it satisfactory and recommend that it be accepted.

______________________________
Mark VanDam, Ph.D., Chair

______________________________
Nancy Potter, Ph.D.

______________________________
Georgina Lynch, M.A.
WEB-BASED TELEREHABILITATION ASSESSMENT
OF RECEPTIVE LANGUAGE

Abstract

by Alyssa Anderson, M.A.
Washington State University
May 2014

Chair: Mark VanDam

Telerehabilitation has the potential to extend services to unreached populations in speech-language pathology. Evidence-based research is needed for various technologies, assessment tools, disorders, ages and populations to give precedence to assessment via telerehabilitation. This study’s purpose was to demonstrate reliability of a unique web-based telerehabilitation method for assessing children’s receptive language abilities. Thirty-four children between the ages of 4 and 13 were given the Peabody Picture Vocabulary Test – Fourth Edition via two methods, traditional paper assessment and a version available as a web page, in a counterbalanced design. The web-based version was accessed in home settings from a variety of computers and web browsers. It administered all items with scanned images and recorded audio before storing the automatically calculated scores on a secure server. The administrator present for the web-based assessment gave no input; they simply clicked on the child’s gestured or verbalized answer. No significant difference was found between the traditional method and the web-based version for time taken, raw score, or standard score. A high level of correlation was found between the two methods for raw scores and standard scores. The results of this study validate the use of a web-based standardized assessment tool for assessment of childhood receptive language via telerehabilitation. This research also gives precedence for further investigation into web-based methods of telerehabilitation assessment in speech-language pathology.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>SECTION</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Assessment in Speech-Language Pathology</td>
<td>1</td>
</tr>
<tr>
<td>Receptive Language Assessment</td>
<td>1</td>
</tr>
<tr>
<td>Telerehabilitation in General</td>
<td>2</td>
</tr>
<tr>
<td>Telerehabilitation Assessment</td>
<td>5</td>
</tr>
<tr>
<td>Research Questions</td>
<td>9</td>
</tr>
<tr>
<td>METHOD</td>
<td>10</td>
</tr>
<tr>
<td>Participants</td>
<td>10</td>
</tr>
<tr>
<td>Materials</td>
<td>11</td>
</tr>
<tr>
<td>Web-Based Test Design</td>
<td>11</td>
</tr>
<tr>
<td>Visual Stimuli</td>
<td>11</td>
</tr>
<tr>
<td>Audio Recordings</td>
<td>12</td>
</tr>
<tr>
<td>Web Interface Design</td>
<td>12</td>
</tr>
<tr>
<td>Traditional Test Administration</td>
<td>13</td>
</tr>
<tr>
<td>Procedures</td>
<td>14</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>16</td>
</tr>
<tr>
<td>RESULTS</td>
<td>16</td>
</tr>
<tr>
<td>Duration</td>
<td>16</td>
</tr>
<tr>
<td>Scores</td>
<td>17</td>
</tr>
</tbody>
</table>
DISCUSSION.................................................................19
  Research Questions..................................................19
  Advantages of Web-Based Test....................................19
  Disadvantages of Web-Based Test.................................22
  Limitations of Present Study .......................................25
  Implications ............................................................25
  Future Directions ....................................................26
  Conclusion ...............................................................28
STATEMENTS...................................................................29
REFERENCES....................................................................30
LIST OF TABLES

1. Means and Standard Deviations for Data Sets .................................................. 16

2. Select Variations in Administration of Standardized Assessments ...................... 21

3. Behaviors Noted During Testing .............................................................................. 24
LIST OF FIGURES

1. Figure 1; Time by Test Type .......................................................... 17

2. Figure 2; Raw Score by Test Type ...................................................... 18

3. Figure 3; Standard Score by Test Type ............................................. 18
Introduction

Assessment in Speech-Language Pathology

Speech-language pathologists (SLPs) and other clinical professionals administer and interpret a wide variety of assessments across medical, rehabilitation, private practice, and educational settings. Some of the goals of clinical assessment include screening, determining eligibility, differential diagnosis, measurement of severity, revealing levels of performance, directing prognostic outlooks, guiding treatment planning and goal writing, and monitoring progress. The information gleaned from a thorough assessment, in combination with best available research and the individual’s preferences, will guide therapy goals and directly influence the quality of therapy administered. The results of an assessment lead to personal, social, and vocational impact on a client’s life (Baker & Chenery, 1999). Therefore, it is critical that assessments be thorough, varied, valid, reliable, and individualized (Shipley & McAfee, 2009). One particularly important class of assessments is language assessments. In the field of speech-language pathology, standardized language assessments are included in the test battery for a majority of age groups and disorders.

Receptive Language Assessment

Language disorders are one of the most frequent developmental concerns seen by primary health care providers (Fagan & Montgomery, 2009) and it is estimated that seven percent of children in preschool and primary school demonstrate significant impairments in language ability (American Speech-Language-Hearing Association [ASHA], 2008). Expressive and receptive language disorders comprise the majority of new referrals to speech-language pathologists (Waite, Theodoros, Russell, & Cahill, 2010b), and, according to the 2010 ASHA Schools survey, 89.9% of speech-language pathologists in the United States have children with language impairments on their caseload. Of those that have clients with language impairments,
the number per caseload averages 24.1 students (ASHA, 2010). Deficits in language abilities may impact children's academic, social, and emotional development (ASHA, 2008).

Receptive language impairments are characterized by difficulties with comprehension of spoken or written language. It is critical that these children have access to clinical services, including speech-language pathology, as soon as possible, and there is strong evidence of the efficacy of early intervention (ASHA, 2008; Guralnick, 2011). It is imperative that reliable and valid assessment and diagnosis be a part of the intervention process (Baker & Chenery, 1999). Poor vocabulary skills often contribute to receptive language impairment and a standardized assessment of receptive vocabulary is routinely included in an assessment battery. Given the lifelong impact that a language disorder in childhood may have on an individual (Beitchman et al., 2001), early identification and thorough assessment are crucial.

**Telerehabilitation in General**

Telerehabilitation is the provision of rehabilitation services via information and communication technologies by professionals such as physical therapists, occupational therapists, speech-language pathologists, audiologists, rehab physicians and nurses, psychologists, teachers, and dieticians, among others. Clinically, telerehabilitation encompasses a broad span of services including assessment, monitoring, prevention, intervention, supervision, education, consultation, and counseling (Brennan et al., 2010). These services are delivered across distance using such modalities as the Internet and/or telephones. ASHA defines telepractice as, “the application of telecommunications technology to the delivery of speech language pathology and audiology professional services at a distance by linking clinician to client/patient or clinician to clinician for assessment, intervention, and/or consultation” (ASHA, 2014). While ASHA consistently uses the word telepractice, the term telerehabilitation is also acceptable and equivalent in its clinical definition, thus the two terms are used interchangeably here. The terms telehealth or telemedicine refer to broader categories of health services.
There are numerous barriers in accessing appropriate speech-language pathology services including: distance, especially for individuals in rural or remote areas, lack of transportation, family work schedules, limited finances, and SLP shortages (Waite et al., 2010b). Also, local speech-language pathologists may not have the expertise to assess, diagnose, and/or treat rare cases. Telepractice could conceivably become one solution to the access problem. Telepractice has the potential to extend services across distances, including those from various underserved populations (ASHA, 2005).

The most frequently utilized model for telerehabilitation service delivery connects the clinician to the client via synchronous (real-time) interactions using two-way audio and video connections. Using high-resolution videoconferencing systems allows for a high level of interactivity between clinician and client. These live video feeds can be supported with web-based applications such as online remediation materials, collaboration tools (e.g., Google Docs, DropBox, etc.) or other telecommunication equipment (fax, e-mail, document scanners, document cameras), which can augment the videoconferencing experience. However, this model depends on sufficient bandwidth and network connection speeds that may not accessible, thus limiting the availability of services (Williams, 2013). There are also asynchronous “store and forward” methods of telerehabilitation in which recordings, videos, or other media are collected, then sent to the clinician for viewing and interpretation (ASHA, 2014). As telerehabilitation has gained acceptance and is becoming increasingly reimbursed by third party payers, access to services and research on telepractice has increased exponentially in recent years (American Telemedicine Association, 2012; Whitten & Love, 2005). New technological developments may further increase the need for additional research. For example, a recent meta-analysis searching for evidence to support the use of Skype, one of the most common telecommunication applications, found no evidence to support or refute the use of this application (Armfield, Gray, & Smith, 2012).
Telerehabilitation has been utilized for patients with a variety of communication and swallowing disorders including: aphasia, dysarthria, apraxia, cognitive-communication disorders, dementia, right hemisphere syndrome, fluency disorders, voice disorders, dysphagia, speech sound disorders, autism, learning disabilities, specific language impairment, language delay, and a variety of other language disabilities. Treatment for a wide variety of pathologies has been pursued, including cerebrovascular accident, traumatic brain injury, Parkinson’s disease, multiple sclerosis, cerebral palsy, head and neck cancer, laryngectomy, ALS, muscular dystrophy, hearing loss, Down syndrome, and cleft palate (Mashima & Doarn, 2008). Studies with these populations have spanned all ages and healthcare settings. In general, telerehabilitation research with these disorders has been cautiously positive, both in reliability/validity and clinician/client satisfaction reports (Mashima & Doarn, 2008). For each of these disorders, accurate assessment is critical to lead to effective therapy; thus research in telerehabilitation assessment for each of these disorders is critical.

Currently, a number of telepractice programs are conducting assessment, treatment, consultation and more with hundreds of patients (Hound & Trottier, 2006). One program had already completed 11,000 speech therapy sessions in the public school system of rural Oklahoma by 2006. The speech-language pathologists worked with a variety of disorders including speech sound disorders, language impairments, autism, fluency disorders, and children with hearing impairments. The district administrators and SLPs reported positive results and high levels of satisfaction with the program (Forducey, 2006). Another study analyzed the data of several SLPs practicing telepractice in the public schools with 71 students, concluding that telepractice is a viable service delivery model for speech-language therapy delivered to public school students (Gabel, Grogan-Johnson, Alvares, Bechstein, & Taylor, 2013). Programs such as these have been shown to be successful and clients and professionals have been satisfied with the platform of telerehabilitation (Crutchley & Campbell, 2010; Theodoros, 2012). This is consistent with the high level of patient and provider satisfaction.
reported for telepractice in other healthcare fields (Whitten & Love, 2005; Whitten & Mair, 2000).

To date, speech-language pathology telerehabilitation research has mainly focused on treatment. This could be due to several factors. First, many standardized tests are not flexible to modifications. Second, assessment requires clinician/client interaction that may not be suitable for telepractice. Also, certain confounding environmental factors may play a different role in live assessment versus telepractice test administration. Lastly, research is inconclusive as to what the best technology options are for real-time assessment of various populations. Studies, while positive in their outlook, vary greatly in their utilized methodologies, from specially constructed set ups and programs to publically available programs such as Skype. In general, computer programs designed for use in speech therapy have tended to focus on treatment rather than assessment (Popovici & Buica-Belciu, 2012), so research on the few specific programs available is sparse.

Telerehabilitation Assessment

In 2005, the American Speech-Language-Hearing Association released a position statement on telepractice stating that the telepractice delivery model must provide the same quality of service that is provided through face-to-face practice. ASHA’s position is that services provided by telepractice must adhere to current policies including the code of ethics, scope of practice and state or federal laws (ASHA, 2014). In order to comply with these guidelines, high quality evidence-based research must demonstrate the reliability and validity of assessment methods with a variety of populations in a telepractice environment. While research has shown many assessment tools and technology models to be effective, the pace of telerehabilitation progress in the field of speech-language pathology is advancing at a rapid rate. In a recent ASHA Leader article, one speech-language pathologist expressed her dismay and struggles in early intervention due to the lack of standardized instruments proven for use in telepractice (Sippl &
Ciccia, 2014). The author also expressed frustration with assessment for younger children in that it is difficult to keep their attention with longer assessments, and that scheduling is difficult between busy families and clinicians, especially across time zones (Sippl & Ciccia, 2014).

Several studies have addressed assessment of adult populations via telerehabilitation in speech-language pathology. One study in 2008 focused on assessment of language disorders in adult populations using the Boston Diagnostic Aphasia Examination (BDAE-3; Goodglass, Kaplan, & Barresi, 2001) and its accompanying Boston Naming Test (BNT, 2nd Edition; Kaplan, Goodglass, & Weintraub, 2001). In that study, 32 participants with aphasia due to stroke or traumatic brain injury were evaluated, 15 on the online assessment and 17 via traditional paper assessment. They also administered a participant satisfaction questionnaire to those assessed via the Internet. A specialized computer was set up on-campus at the University of Queensland that connected to the patient computer at a local hospital. Both computers were touch screen with two webcams: one for live feed, another to record. Results did not reveal a significant difference between the online assessment and standard procedures for any of the subtests. The participants reported high satisfaction with the online testing procedures. The researchers concluded that assessment of aphasia over the Internet is generally reliable and valid and can lead to equivalent diagnosis of type and severity of the disorder (Theodoros, Hill, Russell, Ward, & Wootton, 2008). They did report struggles with the bandwidth resulting in delays of video and audio, which in turn affected the communication between the clinician and the patient. The positive results of this study are consistent with other studies on adult language assessment with aphasia (Georgeadis, Brennan, Barker, & Baron, 2004; Hill, Theodoros, Russell, Ward, & Wootton, 2009b; Palsbo, 2007). Other assessment studies with adults have supported the assessment of apraxia (Hill, Theodoros, Russell, & Ward, 2009a), cochlear implant measures (Hughes et al., 2012), pure-tone hearing testing (Yao, Wang, & Givens, 2009), dysphagia assessment for those with normal to mild cognitive abilities (Ward, Sharma, Burns, Theodoros, & Russell, 2012), and assessment of dysarthria (Constantinescu et al., 2010; Hill et al., 2006).
With respect to language testing with children, one study at the University of Queensland examined telerehabilitation assessment for children with language disorders (Waite et al., 2010b). Similar to the study using the BDAE with adults, they used an on-campus computer setup with a touch screen computer and multiple webcams. They tested 25 language impaired children ages 5;0-9;11; two-thirds were male. They used select subtests from the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). An onsite SLP recorded results simultaneously while the telepractice SLP recorded results. All audio was recorded for assistance in scoring at a later date. They found no significant difference between face-to-face assessment and online assessment. They reported that online assessment provided valid and reliable results. However, they recommended that future research should focus on the child’s interaction within the assessment since some children in their assessment demonstrated difficulties with maintaining attention and concentration, did not request clarification during audio and video breakdowns, or were unintelligible across the audio feed. They also experienced technical difficulties with video and audio breaking down due to bandwidth limitations, the touchscreen not registering certain responses, room lighting washing out materials, and child positioning limiting camera views. Furthermore, they suggested research focus on various settings, such as schools, clinics, and homes and include children under the age of six. Their study was limited by restriction to a university lab with a specialized computer setup. Also, they used two different groups to test the two treatment conditions, in-person and via telepractice, instead of having the same participants tested in each condition.

Another language assessment tool examined in a telepractice platform was the Test of Word Knowledge (TOWK-2; Wiig & Secord, 1992) in 1996 (Wiig, Jones, and Wiig). The study had 30 students with learning disabilities participate, ages 12;0 to 15;11. The test items were converted to an electronic format. On-site examiners administered general test instructions while the computer administered instructions for specific subtests and test items in a written format. Participants responded by either selecting the multiple-choice answer or typing their
response. Subjects served as their own controls and were administered the test via standard procedures either 3 weeks earlier or 3 weeks later in a counterbalanced design. The computer-based test was administered in groups of 7 in a computer lab. Results indicated that the computer administered test was equivalent to standard methods for time taken when using individual times, but when taking group administration into account an average of 55 minutes was saved per student in administration time. There were significant differences found in the means of the overall test scores and all individual subtests except one. The researchers hypothesized that this was due to the increased cognitive load on the computer test, since all the test items were written only instead of being administered in the traditional written and verbal form. Similarities were discovered between the two administration methods when confidence intervals of 90% were applied to the computer-based scores. The students with learning disabilities and the students with ADD/ADHD expressed preference for the computer method of testing.

For children, studies have found certain telepractice assessments in the field of speech-language pathology to be statistically equivalent for speech sound disorders (Grogan-Johnson, Alvares, Rowan, & Creaghead, 2010; Waite, Cahill, Theodoros, Busuttin, & Russell, 2006), children’s literacy (Waite, Theodoros, Russell, & Cahill, 2010a), augmentative and alternative communication (McDougall, Vessoyan, & Duncan, 2012) and receptive/expressive language screening (Ciccia, Whitford, Krumm, & McNeal, 2011), in addition to the aforementioned topic of receptive/expressive language testing. These studies have all found telepractice assessment to be valid and reliable overall.

One study has examined the validity of electronic assessment of childhood language with specific regards to response method. Haaf, Duncan, Skarakis-Doyle, Carew, and Kapitan (1999) examined how specific presentation and response methods affected reliability of assessment. Using the Peabody Picture Vocabulary Test-Revised [PPVT-R] (Dunn & Dunn, 1981), they tested whether or not presentation via a computer screen affected the participants’ scores.
There were 72 participants from 4;0 to 8;11, all typically developing, from homes with English spoken as the main language. Subjects were randomly assigned to one of three test presentation/response methods: standard presentation/standard verbal response, computer presentation/trackball response, and computer presentation/scanning response. The subjects were trained in their response method and were tested in laboratory or home settings. The study found no statistically significant difference in the three presentation/response methods. Thus computer presentation of the *PPVT-R* stimuli was equivalent to the standard paper presentation for testing receptive vocabulary.

Overall, studies for assessment via telerehabilitation have been positive. There are many assessment tools that have been shown to be statistically equivalent to a telerehabilitation form. Further research is still needed to offer evidence within the evidence-based practice service delivery model advocated by ASHA and to set precedence for telerehabilitation assessment in speech-language pathology. Research is desperately needed in more naturalistic settings, as well as with the younger population. Research has also been inconclusive as to what the best parameters might be for assessment using the Internet. The research should question the different types of technology available and the different models such as asynchronous or synchronous. Asynchronous assessment may be able to limit difficulties inherent with a live connection (e.g., lag, delay, video break up, lost connections, echo, video quality, quality of presented visuals, computer requirements). While positive progress has been made in telepractice assessment research, validity should be examined for each technology system, assessment tool, disorder, etiology, age group, and more.

**Research Questions**

The purpose of the present study was to examine the reliability of a web-based telerehabilitation assessment of childhood receptive language abilities via the Internet. This research seeks to provide evidence-based research supporting unique utilization of the Internet
within telepractice and to further telepractice assessment research with children and with an additional assessment tool. The Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4; Dunn & Dunn, 2007) was administered by traditional paper method and compared with a web-based computer application to answer the following questions:

A. Does testing time differ when administered via a web-based computer application as compared with standard administration procedures?

B. Do test scores differ when administered via a web-based computer application as compared with standard administration procedures?

C. Are test time or test scores affected by the external factor of test order?

Method

Participants

Fifty-three children were recruited for the study, but due to technical difficulties with an earlier version of the software, data from 34 children are included in this report. Thirty-four children between 4 and 13 years of age ($M = 9;6$, $SD = 2;2$) from 15 different families participated in the study. One four-year-old, 5 six-year-olds, 4 seven-year-olds, 4 eight-year-olds, 4 nine-year-olds, 4 ten-year-olds, 8 eleven-year-olds, and 4 twelve-year-olds served as participants. There were more male participants ($n = 21/34$) than female participants ($n = 13/34$). According to parent report, none of the children had hearing loss or uncorrected vision impairment. None had interfering medical diagnoses, and only two of the children previously participated in speech therapy. English was the primary language spoken at home for all the children. Of those for whom ethnicity was reported ($n = 29/34$), 90% were white. Maternal education level served as proxy for socioeconomic status (SES; Chen, Matthews, & Boyce, 2002).
In this study, 76% of the mothers possessed a bachelor's degree or higher. This is notably higher than the national average of 29%, indicating that the children sampled are higher SES than the national average (Pew Research Center, 2013). Of the children tested, all were from homes with at least one computer and all had access to the Internet. Participants originated from rural and urban communities within Eastern Washington. Participants responded to an open recruitment call via phone calls, emails, and/or flyers. Informed consent was obtained in writing from the parent or guardian and from the child.

Materials

The Peabody Picture Vocabulary Test, Fourth Edition, is a receptive vocabulary test standardized from preschool through the lifespan. The PPVT-4 was chosen as the experimental measure due to several factors. First, there are two separate but statistically equivalent forms (A and B), thus allowing multiple administrations while limiting carryover effects. Second, the test demands four-alternative forced-choice feedback, ideal for computer administration and scoring. Finally, the PPVT-4 has a long history of administration with extensive standardization and is commonly used by speech-language pathologists and other professionals.

Web-based test design. The traditional paper version of the test was interpreted into a web-based format, accessed via a web browser, allowing for worldwide access from any computer with Internet access.

Visual stimuli. Form A images were provided by the test publisher (Dunn & Dunn, 2007). Form B images were digitally scanned as high quality, color TIFF files. All image files were centered, cropped, and downsampled to 96 dpi and 100 by 760 pixels in JPEG image format. The finished images were visually identical to the panels in the manual. The
bookbinding was not visible in the digital image reproductions, but image numbers, separation lines, and page numbers were retained.

**Audio recordings.** Trained, adult-female voices recorded all stimuli, carrier phrases, and test audio using high quality microphones and digital recording equipment in a sound-attenuated IAC booth at the Speech and Language Lab at Washington State University. Audio was amplitude normalized and digitally mastered to minimize noise and extraneous sounds.

**Web interface design.** Accessed via a hyperlink to a secure website, the tester entered participant details including the child’s name, age and desired test form (A or B). Based on the inputted age, the test automatically selected the appropriate start item. An introductory screen introduced the procedures to the child and presented several practice trials. Children were instructed to either point to the appropriate item or say the number and the examiner would click on the image. Trial item scripts were written and recorded as close to the manual’s recommended verbiage as possible. Once trial items were passed the test would immediately begin at the age-appropriate start item. A voice recording prompted the examinee to “Point to...[item].” One audio recording of “Point to...” was chosen and then grafted to the recordings of each individual item’s name. Once the child selected an image, the examiner clicked directly on the chosen image with a mouse or laptop trackpad. Selecting an image advanced the test to the next panel. Approximately one centimeter of unselectable “dead space” between the four image choices was used to prevent errant selections. For each item, the child was allowed ten seconds, then a second prompt would play saying, “Try one. Point to the one you think it might be.” This second audio prompt followed the manual’s recommendations for time and wording. After the second audio prompt played, another ten seconds was allowed before the item was scored as incorrect and the test proceeded. The child could also state “I don’t know,” or “Pass” on any item and the examiner would click the “I don’t know” button.
Software recorded correctly identified items in real time, and the software implemented adjustments as necessary. For example, the trial start panel reverted to the younger set if certain items were identified incorrectly according to the criteria defined by the test manual. If these criteria were not met, the test would end. The browser’s back button was disabled and the refresh button restarted to the beginning of the examination to ensure that each item was only administered once and no manipulations were made to the progression of the test.

Software automatically calculated raw basal and ceiling scores and would progress backward or forward as needed. Once the ceiling was reached, the test would end and a farewell screen was displayed. Raw scores were automatically converted to standard score, confidence intervals of 90 and 95%, percentile, Normal Curve Equivalent, growth scale value, and stanine by accessing a digital database of look-up tables from the PPVT-4 Manual (Dunn & Dunn, 2007). Tables provided in the paper version of the test were converted to digital format and stored on secure servers for the testing program to access. In addition to all standard measures obtained from the test, software collected testing timing and duration measure, item-level response data, and other variables. These data were stored on a secure server, with each examinee’s individual information linked to test performance data. No computation on the part of the examiner was necessary, and no feedback was provided to the user during or after the test.

**Traditional test administration.** The instruction manual was followed for administration of the paper version. Although the web-based test was designed to mirror the paper method, certain concessions described above were made to accommodate the different administration medium. We restricted some of the optionality in the instructions to make for greater consistency between the paper and web-based formats. The intention of the consistency was so that these scores could be compared to the normative sample of the PPVT-4. Any small deviations from the manual’s recommended procedures are explicitly stated here.
To maintain consistency throughout administrations of test type, the prompt of “Point to...” was used for every item, instead of dropping the carrier phrase once the child became familiar with testing procedures as the manual suggests. This was done for both paper and web-based administrations. The same trial item script was also utilized for both administration types. Due to the restrictions of the web implementation, there was no opportunity for participants to return to a previous answer or ask for a repetition of the audio stimuli. Both changing a previous answer and requesting a repetition of the auditory stimuli were allowed in the paper administration. The manual specifies for examiners to allow the child 10 seconds, then to prompt with the phrase, “Try one. Point to the one you think it might be,” as was used for the web-based administration. The manual then indicates that if the child does not respond to proceed and score the item as incorrect. In the web-based test, this was limited to an additional 10 seconds, but these times were not standardized for the paper administration. If the examiners felt that the child was still thinking, additional time may have been given, or a prompt of, “It’s okay to guess,” another suggestion from the manual, might be given.

In the web-based administration, distractions or conversations brought about by the child could interfere with the child’s ability to hear the next audio stimuli. During the paper administration the examiner would address these conversations but cut them off relatively quickly so as to not drag on administration.

The examiners noted on the paper record form if the child changed an answer, asked for repetitions, started a conversation, or if any outside distractions were noted (ex. the family dog interrupting). The examiners also recorded start and end time for each test. Scoring time was not included in recorded times.

**Procedures**

Trained test administrators administered the tests. A pilot program was conducted with 19 children in order to determine that the examiners were consistent, that the web-based
program functioned properly, and that the scores were being recorded accurately. During this initial pilot testing, inconsistencies were fixed. Examiners consulted with each other and a senior examiner to ensure that administration was consistent and accurate across the types of test administration and with each other.

Parents were allowed to observe testing sessions, but were instructed not to give any instructions or input to the child. Tests were administered in home environments. A quiet room free of audio and visual distractions was arranged in the participants’ homes. Nevertheless, certain distractions from typical home life were observed (e.g. neighbors running various outdoor machinery, children playing games in other rooms, televisions and gaming systems being turned on, construction projects being continued elsewhere in the house, and so on) but were not controlled for.

Examiners' laptops or the home computers of the participants were used for the web-based version, depending on parents' preference and/or ease of access to the wireless Internet. Over the course of testing some of the computers used included: Dell laptop, Dell PC, MacBook Pro, and iMac. Because of minimal technical difficulties with Internet Explorer, browsers Google Chrome, Mozilla Firefox, and Apple Safari were utilized in the administrations of the online examination. No modifications of the computer or browser were necessary, provided Internet access and the sound were both functioning. Audio played from the computer speakers was adjusted to a comfortable listening level determined by the examiner.

Test version (A and B) and test medium (paper and web-based) were pseudo-randomized for even distribution among participants. Due to the different forms, examinees were given the second test directly following the first, similar to the method used during the standardization of the forms for the PPVT-4 (Dunn & Dunn, 2007). Examinees were given small prizes or pieces of candy following each test administration.
Statistical Analysis

Variables included the raw and standard scores for all tests. Scores were coded for format (paper, web-based), test type (Form A, Form B), test time (in minutes), order of test given (first, second), and child age. Demographic variables including child sex, mother’s education, father’s education, zip code, ethnicity, primary language, computer access, Internet access, sensory impairments, and medical diagnoses were not utilized in relation to statistical relationships. The nonparametric Wilcoxon paired sign-rank test (two-tailed) was used to compare variables, since the variables are not in normal distributions. Spearman rank-order correlations were run to further observe patterns among the variables.

Results

Means and standard deviations for test duration (in minutes), raw scores, standard scores for paper administration, web-based administration, and both types of administration are shown in Table One.

Table 1: Means and Standard Deviations for Data Sets

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Paper Mean (SD)</th>
<th>Web-Based Mean (SD)</th>
<th>Combined Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (in minutes)</td>
<td>12.7 (3.5)</td>
<td>11.1 (2.3)</td>
<td>11.9 (3.0)</td>
</tr>
<tr>
<td>Raw Scores</td>
<td>157.7 (26.2)</td>
<td>155.3 (27.2)</td>
<td>156.4 (26.5)</td>
</tr>
<tr>
<td>Standard Scores</td>
<td>111.7 (12.1)</td>
<td>109.6 (11.4)</td>
<td>110.6 (11.7)</td>
</tr>
</tbody>
</table>

Duration

With an n of 28, the Wilcoxon paired sign-rank test did not reveal a significant difference between the administration time of the paper test and the administration time of the web-based test ($p > 0.05$). Due to recording failure on select paper tests, duration was not analyzed for 6 participants. The 28 remaining participants’ times for paper and web-based administrations are
shown in Figure One. Alternative graphic representations of Figures One and Two were presented by Ahmann, Anderson, VanDam, and Potter (2014). Spearman rank-order correlations revealed a moderate positive relationship in times between the two test types \( (r = 0.35) \). Regardless of order of test version administered, as in paper first or web-based first, analysis did not reveal a significant difference \( (p > 0.05) \). No significant difference was revealed for time in comparing all of the tests administered first to all those administered second \( (p > 0.05) \).

![Time by Test Type](image)

**Figure 1**

**Scores**

No significant difference was found between the paper format and web-based administration for raw score or standard score \( (p > 0.05) \). Spearman rank-order correlations revealed a strong positive correlation between the two sets of raw scores \( (r = 0.86) \). Raw scores for both paper and web-based administration for each participant are shown in Figure Two. Spearman rank-order correlations revealed a strong positive correlation between the two sets of standard scores \( (r = 0.74) \). Standard scores for both paper and web-based administration for each participant are shown in Figure Three. Analysis revealed no difference in the raw scores or
standard scores when test types were compared in relation to test order ($p > 0.05$). No significant difference in raw scores or standard scores when comparing all test given first and all those given second ($p > 0.05$).

Figure 2

Raw Score by Test Type

<table>
<thead>
<tr>
<th>Participant</th>
<th>Raw Score by Test Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3

Standard Score by Test Type

<table>
<thead>
<tr>
<th>Participant</th>
<th>Standard Score by Test Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion

Research Questions

This study aimed to determine the reliability of web-based telerehabilitation assessment of children’s receptive language abilities using a standardized assessment tool, the PPVT-4. The results demonstrated the reliability of the web-based administration and scoring for test time, raw score, and standard score. Overall, a strong relationship was observed between the raw scores and standard scores between test administration modalities. Significant difference was not found between overall test duration on either method of administration or between testing order.

Advantages of Web-Based Test

This study uniquely examined fully automatic, web-based standardized test administration in a telerehabilitation service delivery model. Previous research had demonstrated the feasibility of computer-based standardized test administrations. There are many advantages to web-based administration types including access, security, automaticity, overall administration time, and participant engagement.

First and foremost, this allows for telerehabilitation utilization in a variety of settings. The administrator and the child could reside in any location across the world with Internet access. In a recent study, 71.7% of Americans reported having Internet access in their home and 75.6% of households reported having a computer in the home (File, 2013). This gives an optimistic outlook for web-based assessment within telerehabilitation. With the web-based format there are no requirements for the computer to be Windows, Macintosh, Linux, etc., nor does the user have to have a specific program installed. Thus, a computer with Internet access is the only prerequisite. Also, with the asynchronous design of the web-based format, the test could be taken at any time, so frustrations with coordinating between time zones would be eliminated. Therefore, the test can be accessed in any location at any time and it can be accessed
from any computer or laptop with Internet access. No appointments need to be coordinated, no physical materials need to be exchanged, and no one is required to travel to an inconvenient location. This web-based telerehabilitation model is practical for many clinicians and clients and has extensive implications for persons with limited access to specialized services including those with rare disorders or those living in rural areas.

When surveying state administrators across the nation on why telerehabilitation was not a more present delivery model, especially when it could help so many, one of the main concerns reported (40.7%) was security (Cason, Behl, & Ringwalt, 2012). Security is an important and intensely relevant issue in healthcare fields especially when discussing services delivered via the Internet. With the web-based program presented here, test results are only accessible to those with the credentials to access the secure server. Also, all data is transferred in an encrypted, secure format.

Another advantage of the present program is the high degree of automaticity. Because the test presents all stimuli, the administrator needs little familiarity with test administration because their input is minimal. Most importantly, this creates a high level of reliability. Despite the extensive training of speech-language pathologists, there is great variability in administration of standardized assessments, which Maynard and Marlaire (1992) attribute to the “interactional substrate” (p. 193). This interactional substrate is the unique interplay between the child and clinician, which may influence assessment results. It involves behaviors such as prompting, indicating accuracy of answers, correcting responses, or responding to a child’s body language or verbalizations. Given the variety in the interaction during an assessment, administrators are not the neutral conduits we idealize them to be. These variations are not necessarily from inabilities or incompetencies of the examiner or examinee, but are instead related to the distinctive interaction between both parties in a given assessment session (p. 198). The authors suggest that the interactional substrate more closely approximates an environmental factor that has not been adequately examined or controlled for (p. 194).
Positively, it is the scaffolding upon which clinicians gain access to measurable and quantifiable abilities. On the other hand, these behaviors may affect the outcome of the test and the validity of the assessment. And, as Muskett, Body, and Perkins (2012) point out, it is common not to report the specific details of an individual test administration. Results are presented as objective values, with little reference to the specific behaviors and interaction between the clinician and the client that may affect the results of the standardized test. Rarely would a clinician report their own facial expressions in response to certain answers or the exact body language of the child. This practice of not reporting environmental factors in assessment ignores the fact that variations among examiners may impact scores. Administration can vary from child to child, from disorder to disorder, and from day to day. Table Two presents a sample of the variations that may exist.

Table 2

<table>
<thead>
<tr>
<th>Select Variations in Administration of Standardized Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time given for responses</td>
</tr>
<tr>
<td>Accuracy expected for correct responses</td>
</tr>
<tr>
<td>Number of opportunities given for responses</td>
</tr>
<tr>
<td>Speed of administration</td>
</tr>
<tr>
<td>Vocal volume</td>
</tr>
<tr>
<td>Emphasis/Intonation</td>
</tr>
<tr>
<td>Amount of cueing/prompting</td>
</tr>
<tr>
<td>Type of cueing/prompting</td>
</tr>
<tr>
<td>Gestures/Body language</td>
</tr>
<tr>
<td>Facial Expressions</td>
</tr>
<tr>
<td>Engagement in unrelated conversations</td>
</tr>
<tr>
<td>Additional instructions</td>
</tr>
<tr>
<td>Rewards/prizes</td>
</tr>
<tr>
<td>Praise and encouragement</td>
</tr>
<tr>
<td>Knowledge of performance</td>
</tr>
<tr>
<td>Amount of training to take test</td>
</tr>
</tbody>
</table>

Standardized tests are designed to elicit responses dependent solely on the student’s abilities, with no influence from the examiner and his/her administration methods or style. This ideal consistency is taken as an accepted truth (Muskett, Body, & Perkins, 2012, p. 88). When
critically examining this interactional substrate, the variations among assessments are astounding. An automatically administered web-based format can minimize or eliminate many of these variations.

While there is no statistically significant difference in time taken for the tests in the current work, test administrators and participants reported the subjective impression that the web-based administration “felt faster.” This may be due to the consistent 10-second time limit before the first prompt, and the limited additional 10 seconds to respond. It could also be because, after answering, the test immediately started administering the next item. This automatic progression contrasts with waiting for the examiner to turn the booklet, find his/her next stimulus to present, take a breath, and verbalize the next item, all with varying speeds and/or interruptions.

One last anecdotal advantage of the web-based administration is that the children appeared to be more engaged in the web-based computer test. Instead of looking around, making lots of comments, and fidgeting, the examiners noted that when they were given the web-based administration their attention to task increased. They were attentive to the computer and less frequently lost focus. Many seemed excited to interact with the computer and some were even disappointed that they were not given the opportunity to click on the correct answer themselves. With a new generation of children who are more comfortable with and more exposed to technology, computer-based testing methods will continue to have more of a place in healthcare and education fields.

**Disadvantages of Web-Based Test**

Despite the many advantages of a web-based system, there are differences with traditional paper assessment that are still not accounted for. These include limitations on interactions, technical difficulties, and the removal of clinical judgment.
Although a child producing fewer off-task verbalizations child during the computer test is a positive factor, the speed of the web-based administration does limit the interaction between the child and the examiner. In informal observations written during the paper administrations, the examiners noted the number of times a child asked for a repetition, initiated a conversation, or changed an answer. Table Three shows for each observation how many students of the 34 showed the specific behavior at least once, the percentage of the total students this represents, and the average number of times they demonstrated the behavior. Approximately one third of the children requested a repetition, one half initiated a conversation during the assessment, and two thirds changed at least one initially verbalized answer. During web-based testing repetitions, conversations, and changing an answer post-facto were not allowed. The conversations are not necessarily related to test performance, but extra repetitions of the stimuli and the opportunity to change an answer after answering could both affect how well some children perform. As seen in Table Three, these behaviors do occur during testing as a relatively common occurrence, but only once per child, per test, on average. Although an effect was not observed with typically developing children in the present sample, known increased error and difficulty with standardized testing in disordered populations may be more easily influenced by these modifications to traditional testing methods. The lack of time to be off task could be an advantage to keep students with attention deficits on task, but might also impact them if they are not able to keep their attention consistently on the test. They may need breaks even in the short 5-15 minutes the test takes. The present test version does not have the capacity to pause built in as an effort to maintain reliability in administration. Addition of a feature to pause the presentation may be a useful feature to pursue in future research. During testing only one child left the testing room for a restroom break during the test, but breaks such as this present additional need for a pause feature.
### Table 3: Behaviors Noted During Testing

<table>
<thead>
<tr>
<th>Observation</th>
<th>n</th>
<th>Percent of Children</th>
<th>M (Times per child)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitions Requested</td>
<td>10</td>
<td>29%</td>
<td>0.58</td>
</tr>
<tr>
<td>Conversations Initiated</td>
<td>16</td>
<td>47%</td>
<td>1.12</td>
</tr>
<tr>
<td>Answers Changed</td>
<td>23</td>
<td>68%</td>
<td>1.15</td>
</tr>
</tbody>
</table>

With respect to technology, there are always technical difficulties to be dealt with. With the current telepractice application, reliance on the Internet connection could pose a problem. During initial testing, there were early administrations that failed due to a slow Internet connection, but this problem was resolved for the data reported here. Despite best efforts, the Internet may be unpredictable during administration. However, this particular assessment model eliminated the continuous video feed of other telepractice models. There may be additional issues with the interoperability between operating systems, hardware manufacturers, or web browsers. For example some tests showed slightly aberrant behavior of the interface when using Internet Explorer, although this was the only technical difficulty due to browser compatibility observed. This malfunction could be fixed eventually, with more work on the programming end, but it does illustrate that technology presents unique challenges.

For consistency purposes, one difference with the web-based administration was the presentation of each stimulus within the carrier phrase, “Point to...” Since there is no administrator determining when to leave the carrier phrase by default it had to be repeated for each item. This may have caused test administration to take longer and/or could have been distracting to the child.

The present study’s web-based program did not allow for the examiner to use clinical judgment to individualize the assessment. Some children with intellectual, attention, or other disorders, as well as children from linguistically diverse backgrounds, may have needed
additional administrator input to get through the test. The online test could not be individualized to their unique needs. The administrator was stripped of the opportunity to use judgment if, for instance, the child seemed to need a break, if they answered without thinking, or if they were answering with a consistent pattern. These and so many more behaviors are seen and addressed by clinicians in the assessment process and become another observation of the child’s abilities. Also, some research has shown that a more dynamic testing method may be beneficial (Muskett, Body, & Perkins, 2012). Additionally, if the clinician was not present they might be unable to note their clinical observations during the assessment. However, the approximate 15-minute duration of administering this specific assessment tool is just one part of a well-rounded and thorough language assessment.

Limitations of the Present Study

Because of the recruitment methodology, the test population was relatively similar across demographic factors. Testing children all from a higher socioeconomic status led to a higher than average set of test results. None of the children tested were more than 1 Standard Deviation below the mean, meaning that none of the children tested would fall into the disordered or at-risk categories. No children with documented disabilities were represented in the test population either. The last limitation is that while the examiners made every effort to be as uninvolved as possible for the web-based administrations, their experience with standardized testing may have led to some bias or extra input during testing.

Implications

These test results lead to several possible practical implications for clinicians. Many standardized assessment companies are beginning to release electronic versions of their stimuli, including administration and scoring. While more testing is necessary, this research supports the use of these versions in practice. Computerized versions allow the clinician less materials
and a more engaging presentation for the children. As telepractice is on the rise, especially in the field of speech-language pathology, this research gives precedence to assessment within a telerehabilitation environment and its efficacy. This research especially emphasizes the benefits of a web-based platform and its possibilities for standardized testing. Lastly, this research allows examiners to look to the future for new utilizations of the Internet within their practice.

**Future Directions**

With the many implications of this research come many more questions. These questions can be expanded in many directions for future research, which include varying administrators, populations, tests, methods of interaction with the computer program and more.

Further research in this area should focus on allowing persons with no administration experience to help the children take the test. This could include parents, para-educators, speech-language assistants or other clinic/school staff who may be available to assist with administration of the test. The main advantage of the web-based test is its ability to be taken in any location. Therefore, it would build on this research for the test to be able to be taken by anyone without the presence of a qualified administrator. This is not to diminish or eliminate the role of clinicians in the assessment process, whom remain responsible for choosing the test battery tools, the interpretation, and the application of the results. Clinicians are still tasked with the understanding and synthesizing clinical data and making judgments about the validity of this data (Cochran, 2005, p.299). During interpretation of assessment data the clinical judgment of clinicians is the most crucial (Cochran, 2005, p.344). Instead of diminishing the role of a clinician, a web-based assessment can help to speed up the assessment process for busy clinicians, or, more importantly, allow children access to testing who might not have had access to standardized assessments.

Future research may also focus on administration with other populations such as: low SES, linguistically diverse, autism, intellectually disabled, attention deficit disorder, language
disordered, and others with disabilities. Because many disabilities occur with physical or sensory impairment, the Internet platform for assessment may allow even more options for modifications with these populations such as switch access or direct connections to hearing aids.

Since this research has demonstrated that one test is feasible in a telerehabilitation platform, research should continue with other tests that could be converted to a web-based format. As previously discussed, research has started to appear focusing on testing via telepractice, but the researched materials have not been web-based in that the materials are accessed on a website. Other tests provide more unique challenges when they require verbal input from the child or more complex stimuli presentation. They might require more back and forth with the clinician or physical manipulatives. Variables to consider include videoconferencing program (e.g., Skype, Oovoo, Google Hangout, etc.), bandwidth, transferring of materials, type of computer, type and quality of webcam, and ease of use for patients. While these present distinct challenges, the constant innovation in technology leads to a positive outlook for these possibilities.

Billing in healthcare fields is a multi-faceted and complicated issue, yet it is more relevant than ever with changing insurance guidelines and more. The cost in this study is not readily observed due to the research setting. However, for each web-based test given, the cost for each test administration may be reduced, such as with the removal of printed test protocols. Other aspects of cost, such as billable time for the clinician with testing and scoring could also yield potential savings for clients and clinicians. The cost effectiveness of the telerehabilitation platform, especially with regards to asynchronous assessment, lends itself to further investigation.

The examiners chose to click the stimuli themselves for accuracy and consistency. Some of the older children may have been able to accurately click on their own answer. This depends on their experience with computers, their accuracy with a mouse or trackpad, and the individual interplay between the desire to perform well and the desire for the test to be finished. It raises
the question of the appropriate age for children to be able to select for themselves and could add another degree of variability. Other possibilities include touchscreen computers or tablets, as well as scanning, switches and other access methods.

Lastly, results such as in this study should be positively advocated to the greater speech-language pathology community. Many believe that face-to-face in-person therapy is the “gold standard.” Some are also afraid that the introduction of therapy methods partially or completely conducted via technology may diminish the need for speech-language pathologists. Instead of focusing on what the SLP would not be doing, the focus should shift to what telerehabilitation can do for both the clinician and the client. Evidence has emerged in an early form that telepractice has unique ways of enhancing quality of life for patients and that the benefits, such as therapy in a naturalistic setting, may outweigh the negatives of telepractice. These naturalistic settings have not only been shown to be more effective, but they are often preferred by clients (McCue, Fairman, Pramuka, 2010). As the future of this profession unfolds, the new generation of SLPs will not be afraid of implementing technology, but they will need guidance and the expertise of experienced clinicians in order to appropriately and effectively implement the technology in their practice (Theodoros, 2011).

Conclusion

The current study revealed no significant difference in time, raw score, or standard score when comparing a web-based administration of the PPVT-4 to the standard testing procedures. Furthermore, agreement was very high in raw score and standard score within participants. Therefore, this study gives precedence to telerehabilitation assessment of receptive language using a web-based version of the PPVT-4 and suggests that other forms of web-based assessment may be viable. It also serves as the backbone for future research with disordered populations, other assessments tools, and with parents administering web-based tools.
**Statements**

This study was reviewed and approved by the Washington State University Institutional Review Board. No conflicts of interest exist for the researchers. Copyright permission was granted by the publishers of Pearson Assessments for use of the test images in a telerehabilitation system. Gratitude is expressed to Derek Blohowiak and Dustin Walker for their continued effort in the programming of the PPVT web-based format.
References


