The Effect of Videoconference-Based Telerehabilitation on Story Retelling Performance by Brain-Injured Subjects and Its Implications for Remote Speech-Language Therapy

DAVID M. BRENNAN, M.B.E.,1 AMY C. GEORGEADIS, M.S., CCC-SLP,2 CHRISTINE R. BARON, M.A., CCC-SLP,2 and LINSEY M. BARKER, B.S.1

ABSTRACT

This paper presents results from a study conducted at the Rehabilitation Engineering Research Center (RERC) on Telerehabilitation at the National Rehabilitation Hospital. The study was designed to measure performance by brain-injured subjects, with medical diagnoses of stroke or traumatic brain injury, on a standardized Speech-Language Pathology evaluation conducted in both face-to-face and videoconference-based telerehabilitation settings. The Story Retelling Procedure (SRP), which measures connected language production and comprehension of spoken narratives, was administered to each subject in both settings. The primary objectives of this study were to: (1) compare communication as measured by the SRP between experimental settings, and (2) determine if subject variables (such as age, education, technology experience or gender) had an effect on performance differences between settings. The rationale was that any difference in this aspect of performance must be identified and characterized before this mode of intervention can be used clinically. Across all subjects \( n = 40 \), no significant difference \( p > 0.05 \) was found between SRP performance measured in the two settings. Additionally, variables including age, education, technology experience, and gender did not significantly affect the difference between performance in the two settings. Overall, subjects reported a high level of acceptance of videoconferencing with 34 subjects responding “yes,” 4 responding “no,” and 2 responding “maybe” when asked if they would use videoconferencing again to talk to a clinician. Results of this study confirm the potential for SLP treatment using videoconferencing and indicate a need for continued research in the field.

INTRODUCTION

Using modern communication tools and techniques, the traditional rehabilitation service delivery model is being expanded. Home accommodation and accessibility assessments, vocational rehabilitation counseling, and wound care management are just a few examples of telerehab applications where services can be provided remotely to areas and clients where these services would not otherwise be easily accessible. The concept of providing remote speech-language pathology (SLP) services to improve treatment access and

1Rehabilitation Engineering Research Center on Telerehabilitation, National Rehabilitation Hospital, Washington, DC.
2Speech-Language Pathology Service, National Rehabilitation Hospital, Washington, DC.
frequency is particularly promising and has been acknowledged since the mid-1970s. One of the earliest efforts in SLP Telehealth made use of basic telephone to deliver services (ranging from client-clinician follow-up for progress review and encouragement to telephone counseling of clients and their families) to veterans with communicative disorders such as a laryngeal speech, articulation defects, stuttering, dysphasia, dysarthria, and auditory and voice disorders.

As technology improved, the types and ranges of services that can be provided to remote clients have expanded. One example of this is videoconferencing, which extends traditional telephone equipment by allowing participants to see (and hear) each other during a call. Videoconferencing is a tool well suited for providing the auditory, verbal, and visual interaction between a clinician and client that is essential in both SLP diagnostic and therapeutic sessions. Research has demonstrated that videoconference-based telerehabilitation is feasible, effective, and appropriate for delivering particular SLP services at a distance. Agreement between diagnoses determined under face-to-face and remote conditions has shown it to be a viable method for providing diagnostic SLP services. To be effective, however, the technology and procedures used during the telerehabilitation session must not distort or interfere with communication, or must do so only in a minimal and well-characterized way.

No published research prior to this study had investigated the effects of videoconferencing on communication between clinician and client, or whether client characteristics such as age, experience, or gender may predict the presence and extent of differences between clients' performance during telerehabilitation and traditional face-to-face sessions. This is important to consider because if, for example, a client is significantly apprehensive or uncomfortable using a new technology such as videoconferencing, it might be reasonable to anticipate that her/his communication will be adversely affected and changes may occur with respect to social pragmatics (e.g., eye contact, turn taking, topic maintaining, etc.), linguistic variables (e.g., language production and content, speed of discourse, etc.), or in other less-predictable ways. The undesirable effects, such as discomfort, anxiety, satisfaction, level of interest, etc., which could arise from technophobia or other client variables, must be considered if a wide spectrum of clients is to be served by telerehabilitation.

This study was designed to compare performance by subjects with brain injuries on a standardized SLP evaluation in two settings, traditional face-to-face and videoconference-based telerehab. The evaluation used was the Story Retelling Procedure (SRP), a picture-supported story-retelling task that measures connected language production by requiring subjects to retain elements of a story and reformulate them in their own words. The study further aimed to determine if descriptive subject variables played a significant role in subjects' performance. Specifically, this study was designed to answer two questions:

1. Does communication as measured by the SRP differ between face-to-face and videoconference-based telerehabilitation settings?
2. Are performance differences between settings, if observed, predicted by age, level of education, prior experience with technology, or gender?

METHODS

Participants

Data were collected from a convenience sample of 44 subjects with varying levels of cognitive-communicative, language, and/or speech impairments resulting from brain injury recruited from the inpatient and outpatient populations of the Speech-Language Pathology Service at the National Rehabilitation Hospital. The subjects all had medical diagnoses of traumatic brain injury (TBI), right-hemisphere cerebrovascular accident (R-CVA), or left-hemisphere cerebrovascular accident (L-CVA). All subjects were less than 1-year post-onset. The Boston Diagnostic Aphasia Examination—Complex Ideational Material subtest (BDAE-CIM) was administered to each subject as a pre-
screening. This subtest serves as a quantitative measure of auditory comprehension and assesses the subject's ability to understand and express verbal agreement and disagreement. To be admitted to the study, a minimum BDAE-CIM score of 8 points (out of a possible 12 points) was required. Due to the nature of the SLP evaluation used in the study, additional exclusionary criteria were used to ensure that subjects would be able to participate actively in the testing. These criteria ruled out subjects with documented visual perception deficits such as visual-spatial neglect, field cuts, or similar conditions, as well as those subjects who were not able to engage in simple conversation due to decreased attention or poor speech intelligibility.

Two subjects chose to voluntarily withdraw before testing was completed. During scoring, data from two other subjects were discarded when they were found to be incomplete as a result of equipment problems. Analysis was completed on data from the remaining 40 subjects. This group consisted of 23 males and 17 females ranging in age from 18 to 70 years (M = 43.4, SD = 15.9). Additional subject characteristics are summarized and presented in Table 1.

MATERIALS

One standardized method for evaluating a subject's communication skills is to assess her/his comprehension and retention of spoken narrative discourse. Doyle et al. built upon earlier work by developing a picture-supported SRP. The SRP measures connected language production by requiring subjects to retain the elements of a story and reformulate them linguistically. In the SRP, the subject listens to a prerecorded story accompanied by individual pictures that are shown in series to correspond with events in the story. At the conclusion of the story, all of the pictures are displayed together and the subject is asked to retell the study using her/his own words.

The SRP has been shown to be a reliable and standardized procedure for sampling connected language discourse. McNeil et al. developed the Percent Information Unit (%IU) scoring metric as a method for evaluating a subject's performance on the SRP by counting the Information Units (IUs) contained within it. An IU is defined as "an identified word, phrase, or acceptable alternative from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story" (ref. 12, p. 994). The primary benefit to using the %IU metric is a predetermined check-box score sheet for each story, listing all of the IUs in the story and acceptable alternate words or phrases, which makes the process of scoring the stories standardized and repeatable.

The %IU metric has been demonstrated to be a reliable, valid, and efficient method for measuring comprehension and production of spoken narrative discourse. It is highly correlated with other accepted measures of connected spoken language information content such as the Correct Information Unit and the Accurate

<table>
<thead>
<tr>
<th>Table 1. Characteristics of the TBI, LCVA, and RCVA Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Education (years)</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>BDAE-CIM Score (out of 12)</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>
and Complete Story Propositions metrics.\textsuperscript{12–14} McNeil et al.\textsuperscript{15} also validated earlier findings\textsuperscript{11} of "parallel" SRP story sets (each consisting of three stories). Subjects with aphasia tested with sets of parallel stories showed small standard deviations of %IU metric these sets. Their equivalence allows for repeated administration of the SRP using different but equivalent story sets for test-retest purposes.

**PROCEDURE AND EQUIPMENT**

Prior to the start of testing, each subject completed the Technology Experience Survey, a nonstandardized tool developed by the research team for this study. The survey was conducted in interview format with the clinician asking questions to collect information on each subject's overall experience with computers, amount of daily computer usage, and familiarity with different types of computer applications.

Testing occurred in two settings: face-to-face (FF) and remote telerehab (T). Two of the parallel SRP story sets were randomly selected for use in the study. The three stories from one story set (form A) were completed by all subjects in the FF setting, while the three stories from the other story set (form D) were completed in the T setting. To negate practice and test sequence effects, the assignment of test setting order (FF followed by T or T followed by FF) was randomized across subjects. The story text, accompanying pictures, and digital recordings of the stories administered during testing [WAVE files in PCM Format (22.050 kHz, 16 bit, Mono) at 43 kb/sec] are those developed and described by Doyle et al.\textsuperscript{9} and were used with full consent and cooperation.

The FF setting was structured as a conventional therapy session, with the clinician and subject seated together in the Testing Room. This room was a dedicated space used throughout the course of the testing. Its furnishings, size, and noise isolation were consistent with other clinical treatment locations throughout the hospital and were kept constant within and across all subjects. During testing, the clinician controlled the playback of the story sound files and recording of the subject's SRP responses using a computer placed out of view of the subject. While the stories were played, the clinician manually placed full-page prints of the pictures that corresponded temporally with the information in the story on a bookstand in front of the subject. Desktop computer speakers and a noise-canceling computer microphone were placed on the table in front of the subject and used to playback and record the stories.

In the T setting, the subject was located in the testing room, while the clinician was seated in a separate room within the hospital. Because the intent of the study was to test the effects of the intervening telerehabilitation connection at the best available audio and video quality, interaction between subject and clinician was mediated via computer-based videoconferencing with full-duplex audio and video over a high bandwidth (10 Mbps) Local Area Network (LAN) connection. The subject was seated in front of a flat-panel computer monitor placed in the same position and orientation as the bookstand used in the FF setting. Live video of the clinician was displayed on the subject's computer monitor in a "window," approximately 2½ by 2 inches in size, alongside a full-size display of the story pictures (Fig. 1). Similarly, video of the subject was displayed on the computer monitor at the clinician's location. The speakers and microphone were arranged with the same configuration and placement as in the FF setting to play the prerecorded stories and record the subjects' responses, and to allow the subject and clinician to speak to and hear each other. In both settings, the SRP responses were recorded and saved as WAVE files in PCM format (11.025 kHz, 8 bit, Mono) at 10 kb/second.

![FIG. 1. Subject's screen during story playback.](image-url)
The software used during trials in the T setting was developed by the research staff to combine the videoconferencing capability of NetMeeting (Microsoft Corporation, Redmond, WA) with the ability to administer the SRP. The subject was not required to interact with the computer, as the software allowed the clinician to initiate and manage the videoconference as well as remotely control the computer at the subject’s location to present the stories and accompanying pictures, and record the subject’s SRP responses.

At the conclusion of the testing, both the clinician and the subject completed exit surveys to collect their reactions and opinions regarding the session. The Clinician Exit Survey gauged the clinician’s impression of the subject’s communication and interaction with the clinician in both settings. The clinician “interviewed” the subject face-to-face to complete the Subject Exit Survey, which asked the subject to self-report her/his own impression of communication in the FF and T settings. In addition, the survey also asked questions related to the subject’s satisfaction and willingness to use telerehab services.

The time required for a complete testing session (prescreening, technology survey, three SRPs in each setting, and post-test exit surveys) averaged approximately 90 minutes for each subject.

DATA ANALYSIS

The subject’s responses from the stories in both settings were individually replayed and scored by the clinician using McNeil’s %IU scoring metric. For each setting, the %IU scores from the three stories were averaged to arrive at a final score for the subject’s performance in that setting. These scores, FF score and T score (indicating performance in the FF and T settings, respectively), were used to analyze and compare performance across all subjects and settings. A paired samples t-test across all subjects’ FF scores and T scores was used to evaluate the null hypothesis that there is no significant difference in scores between settings.

Due to the objective nature of both the SRP and %IU scoring procedure, bias was not expected to play a significant role in scoring. However, as a single clinician conducted the testing in each setting and scored all SRP responses, the possibility of bias for or against one setting over the other did exist. Interrater reliability was examined to verify nonbias. A second clinician trained in %IU scoring and blinded to the SRP setting scored a random sample of 12 SRP responses (5% of the total number of SRP responses). Percent agreement between the clinicians on the 12 stories was 92.8% (range = 81.1–100%), suggesting a low likelihood of scorer-bias.

To examine the effect of descriptive variables on subjects’ T score versus FF score differences, the 40 subjects were divided into groups relative to each variable:

- Age. The age range of the subjects (18–70) was divided into three equal intervals. Subjects were grouped according to age in the following ranges: 18–35, 36–52, and 53–70 years.
- Education. Subjects were divided, based on highest degree completed, into these groups: less than high school, high school, and college or higher.
- Technology experience. Results from the Technology Experience Survey were used to arrive at a technology experience score for each subject. This score reflected the subject’s prior use of and exposure to computers and was based on the years of computer experience, amount of daily computer use, and the number and types of computer applications used most frequently. The calculated scores ranged from 0 to 24. On the basis of this range, subjects were divided into four groups: Zero (a score of 0), Novice (a score between 1 and 8), Intermediate (a score between 9 and 16), and Advanced (a score between 17 and 24).
- Gender. Male and female.
- To determine the role that each of these independent variables played in performance differences between settings, an analysis of variance (ANOVA) was computed on the difference between each subject’s T score and FF score. This number ($\Delta_{T-FF}$) takes on positive values for higher performance in the
RESULTS

Across all subjects, T scores (the average of each subject’s %IU scores on the three stories in the T setting) ranged from 9 to 70.6 with a mean of 35.1 (SD = 14.6, SEM = 2.3). The range of FF scores was 10.9 to 62.9 with a mean of 34.5 (SD = 13.2, SEM = 2.08). These data are summarized in Figure 2.

A two-tailed paired samples t-test indicated no significant difference [t(39) = 0.69, p = 0.49] and a high correlation (r = 0.93) between T score and FF score across all subjects. Additional data related to the difference in scores between settings (ΔT-FF) across all subjects are shown in Table 2.

No significant differences with regard to ΔT-FF were found among groups for any of the descriptive variables (Table 3). Group differences related to age come closest to statistical significance. The youngest age group (subjects less than 36 years old) performed, on average, better in the FF setting whereas the other two age groups both performed better in the T setting.

DISCUSSION

This study was designed to compare results from an SLP evaluation conducted in traditional face-to-face and videoconference-based telerehabilitation settings. No statistically significant difference was found in performance between settings, indicating that high-quality videoconferencing had no discernible effect on SRP performance by brain-injured subjects.

This study also looked to determine if descriptive subject variables (age, education, technology, experience, and gender) affected inter-setting performance differences. There were no significant differences in performance by groups related to any of the four variables between the T and FF settings. This finding suggests there are no gross exclusionary factors for candidacy for SLP telerehabilitation treatment for brain-injured subjects. It also points to the importance of having an individualized treatment plan for each client with regard to the use of telerehabilitation. A clinician should make decisions collaboratively with each client regarding options for receiving treatment at a distance.

Although not statistically significant, findings regarding effect of age should be noted. Age is sometimes thought to play a large role in users’ attitudes toward technology; however, research has shown this role to be considerably less significant than believed.16,17 Additionally, technophobia and apprehension about technology have been found to exist more or less uniformly across gender, ethnic, and age groups, and to be most dependent on users’ direct experience with technology.18 From those published findings, it was anticipated that the younger subjects in the study would likely have had more previous experience with technology, and that this could result in higher scores in the T setting due to a higher level of acceptance and comfort with the technology.

In actuality, observed differences in performance by age group contradicted this anticipated effect. The youngest age group (18–35 years) was the only group to score, on average, lower in the T setting than in the FF setting.
Table 3. ANOVA Results by Subject Descriptive Variable

<table>
<thead>
<tr>
<th></th>
<th>Mean $\Delta_{T, FF}$</th>
<th>SD</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 36 ($n = 13$)</td>
<td>-1.95</td>
<td>3.32</td>
<td>2</td>
<td>2.26</td>
<td>0.12</td>
</tr>
<tr>
<td>36-52 ($n = 13$)</td>
<td>1.57</td>
<td>5.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater than 52 ($n = 14$)</td>
<td>2.05</td>
<td>6.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (highest degree completed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No high school ($n = 9$)</td>
<td>1.02</td>
<td>7.01</td>
<td>2</td>
<td>0.24</td>
<td>0.79</td>
</tr>
<tr>
<td>High school ($n = 17$)</td>
<td>1.05</td>
<td>4.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College ($n = 14$)</td>
<td>-0.23</td>
<td>6.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero ($n = 8$)</td>
<td>1.98</td>
<td>6.53</td>
<td>3</td>
<td>0.42</td>
<td>0.74</td>
</tr>
<tr>
<td>Novice ($n = 14$)</td>
<td>0.98</td>
<td>4.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate ($n = 8$)</td>
<td>0.35</td>
<td>4.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced ($n = 10$)</td>
<td>-0.86</td>
<td>6.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male ($n = 23$)</td>
<td>1.07</td>
<td>5.00</td>
<td>1</td>
<td>0.41</td>
<td>0.53</td>
</tr>
<tr>
<td>Female ($n = 17$)</td>
<td>-0.05</td>
<td>6.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aNot significant at the 0.05% confidence level.
Positive $\Delta_{T, FF}$ values indicate higher performance in T setting and negative values indicate higher performance in FF setting; a $\Delta_{T, FF}$ value of zero indicates equivalent performance across both settings.

Further, the subjects in the oldest age range (53-70 years), on average, showed the greater superiority of performance in the T setting relative to the FF than either of the two younger groups. One possible explanation for this finding is the confound between age and medical diagnosis. The youngest age group was comprised primarily of subjects with the diagnosis of TBI (9 of 13 subjects). Individuals with this diagnosis, relative to individuals with CVA, report long-term dysfunction due to attention impairments, regardless of the severity of the brain injury.19,20 We hypothesize that the subjects with TBI diagnoses may, in general, have had more difficult time attending to the task in the presence of videoconferencing than when interaction occurred in the FF setting.

It is important to note that the environmental setting in this study was highly controlled. Testing rooms where the subject and clinician were seated were quiet and without external interruptions. The intent was to test the effects of the intervening telerehabilitation connection at the best-available audio and video quality. It is thought that interaction conducted remotely in a less controlled setting (i.e., to a client’s home) has the potential for greater distractions and technical difficulties. This study was also conducted with a group of subjects who had adequate attention and auditory comprehension for a story retelling procedure, good speech intelligibility, and adequate visual acuity. Future studies should address explicitly the impact of the quality of the audio and video signals on videoconference-based SLP telerehab sessions, and also the candidacy for SLP telerehabilitation treatment of clients with different types of medical and SLP diagnoses.

This study employed a diagnostic evaluation designed to measure connected spoken-language production—a specific, valid, and important aspect of communication. In the T setting, subjects interacted with the clinician using videoconferencing for less than 30 minutes and were not required to operate the computer. Future longitudinal studies will build on this work by investigating the impact of a period of SLP therapy delivered via telerehabilitation on the functional communication skills of adult subjects with neurologic impairments. Treatment sessions will combine videoconferencing communication with the ability for the clinician and subject to view and interact simultaneously with programs on the computer, exchange electronic documents, and work together on computerized drill-and-practice software. During these sessions, the subject will have expanded interaction with the computer by using a touchscreen interface, controlling the mouse, and/or typing with a keyboard.

An in-depth investigation of subjective feedback collected from the subjects and the cli-
cian will be presented in subsequent publications. However, a preliminary analysis of the subjects' response to the use of videoconferencing for the provision of SLP services at a distance shows a high level of acceptance. When asked whether they would use videoconferencing again to talk to a clinician, 34 subjects responded “yes,” 4 responded “no,” and 2 responded “maybe.” This positive response, in addition to a lack of significant difference in performance between face-to-face and videoconference settings, illustrates the continued potential for SLP treatment using videoconferencing and suggests a need for continued research and development in the field.

ACKNOWLEDGMENTS

This project was supported by the National Rehabilitation Hospital's Rehabilitation Engineering Research Center on Telerehabilitation funded by the US Department of Education under award number H133E990007.

REFERENCES


Address reprint requests to:
David M. Brennan, M.B.E.
Rehabilitation Engineering Research Center on Telerehabilitation
National Rehabilitation Hospital
102 Irving Street, N.W.
Washington, D.C. 20010
E-mail: david.m.brennan@medstar.net