Interrater Reliability of the Technology Team Assessment Process

Amy J. Embleton, Karen C. Spencer

Key Words: evaluation process, occupational therapy • reliability of tests • self-help devices

Objective. This study examined the interrater reliability of the Technology Team Assessment Process (TTAP).

Method. Students ages 9 to 12 years who have learning disabilities received a TTAP assessment, which was videotaped. Five independent raters trained in the assessment process observed the videotapes and rated each student with the TTAP recording forms. Additionally, each rater made specific assistive technology recommendations on the basis of his or her evaluation.

Results. Intraclass correlation coefficients revealed that only one of four sections of the TTAP investigated had low to moderate interrater reliability for the students evaluated.

Conclusion. The results of this study suggest that the TTAP, in its present form, cannot be used with confidence to assess students with learning disabilities and requires more extensive development and testing.

There is a growing body of literature suggesting that adapted computers or other types of assistive technologies can promote positive changes in the educational performance of children with disabilities (Cook & Hussey, 1995; Hutinger, 1993, 1994; National Council on Disability, 1993; Polloway & Patton, 1993). These changes include increased communication, an enhanced ability to learn, and increased control over one's environment. Assistive technology may also enhance social interaction between students with disabilities and students without disabilities by allowing students with disabilities to participate more fully in a range of educational environments (Male, 1994).

Despite the positive benefits associated with assistive technology, assessments to determine what specific technology a given student needs remains a challenge. Assessments require consideration of the potential user's unique characteristics (e.g., age, mobility, vision); the context for technology use; and, most importantly, what activities the student wants or needs to perform. Ultimately, the activity defines the goal for any type of assistive technology assessment (Cook & Hussey, 1995). For example, a student may need to complete written assignments for an English class. This same student may also need to leave short written messages for family members. The activity focus for assistive technology in both situations is writing. The user, however, may choose different technology devices for each performance environment on the basis of the two different activity demands. Assistive technology assessments that do not consider a person's desire to perform specific activities within real contexts will most likely lead to poor technology recommendations and underutilization of the technology device (DeRuiter, 1995).

The ability to make the best assistive technology recommendations depends on the quality of the assessment and strategies used. Of the few comprehensive tools available to evaluate the assistive technology needs of students with disabilities in educational settings, the Technology Team Assessment Process (TTAP) (Hutinger, 1993) has many favorable characteristics that researchers and authors with expertise in assistive technology and assessment deem important (Angelo, 1992; Bain, 1995; Cook & Hussey, 1995; Holder-Brown & Parette, 1992; Hutinger & Gentry, 1986; Hutinger, Johanson, Robinson, & Schneider, 1992; Mann & Lane, 1991; Parette, Hourcade, & Van-Biervliet, 1993; Rainforth, York, & Macdonald, 1992; Reed & Bowser, 1991). These characteristics include (a) assessing the student in a relevant school, home, or community environment; (b) using meaningful activities versus the technology device to provide the student with a...
focus during the assessment process; (c) formally involving the student and his or her parents, teachers, and therapists during assessment planning and subsequent decision making; and (d) using systematic trials to compare and contrast alternative devices.

The TTAP was developed to address the need for a cost-effective and educationally relevant technology assessment for young students ages birth to 8 years with moderate to severe disabilities (Hutinger, 1993). The assessment has an overall interrater reliability of 98% on the basis of a percentage agreement method using 10 raters. It also has reported face and content validity. Hutinger (1993) further suggested that the TTAP may be applicable to a broader school-age population; however, more extensive reliability and validity testing is needed.

Because of the high incidence of learning disabilities among students who receive special education services and the increasing interest in the use of assistive technologies to mitigate their educational performance deficits, this study examined the interrater reliability of the TTAP for 9-year-old to 12-year-old students with considerable learning disabilities. If found reliable for this age group, the TTAP could prove useful to school-based therapists responsible for assistive technology assessment, device recommendations, and assistive technology services.

The Technology Team Assessment Process

The TTAP uses structured observations of a student’s characteristics and behaviors as he or she performs activities while using assistive technology. Assessment administration requires two teams: a core team of persons with training and expertise in assistive technology and a support team of family members and others who can offer information about the student’s abilities and appropriate goals. The TTAP developers recommend that core team members have extensive expertise with assistive technology to allow them to accurately follow the TTAP protocol and take a lead role during evaluation.

Before administering the assessment, core team members are responsible for reviewing written background information about the student as well as a short videotape of the student engaged in home or school activities. On the basis of these preparatory activities, core team members design or select specific assessment activities and the accompanying software; hardware; adaptations; positioning alternatives; and peripheral devices, such as switches, keyboards, and other items.

The TTAP may be conducted in the home, classroom, or other setting that is familiar and relevant to the student. On the day of the TTAP, core and support team members meet to review the student’s goals, the activity agenda and technology options to be used during the assessment administration, and the two forms on which data are recorded. The first of these forms is the Tech Access Form on which is recorded observed student performance with different input methods. The student’s performance characteristics and behaviors are scored as present (1) or absent (0). Team members also use this form to record specific recommendations for assistive technology devices or services. The second form is the Individual Trial Form, which is used to describe each assessment activity and the student’s response to the activity. The TTAP ends when all planned assessment activities are completed, data are recorded on the appropriate forms, and a report of recommendations for specific technology devices is written.

Given the promising interrater reliability of the TTAP with students ages birth to 8 years (Hutinger, 1993), this study evaluated a broader application of the TTAP by examining its interrater reliability with students ages 9 to 12 years who have learning disabilities. Of interest was the consistency of item scoring on the Tech Access Form across five independent raters as well as the consistency of rater recommendations for specific assistive technology devices.

Method

Participants

Five graduate students in occupational therapy participated as raters in this study. All had received intensive training in the use and application of assistive technology in public school and community settings. Two raters had 1.5 years of training with assistive technology, and three had 5 months of training. All five raters were fully informed as to the purpose and procedures of this study, and each signed a consent form.

Procedure for TTAP Assessment Administration

Three male students who had been recommended for technology assessments by school professionals in a small urban school district were selected for the TTAP. Each student met the following criteria: (a) was between the ages of 9 and 12 years; (b) had a documented limitation in physical, sensory, cognitive, mobility, or communication functioning or any combination of these areas; (c) was not a proficient user of assistive technology devices; and (d) had a discrepancy between his ability and educational performance, indicating that he was performing below his perceived potential.

The students had similar cognitive and motor skills. At the time of the assessment, B. and D. were 9 years, 8 months old, and M. was 12 years, 2 months old. All three students had learning disabilities and fine motor impairments and were receiving special education services. Assistive technology assessments to evaluate the possibility of using a computer for writing were recommended for all three students. Although they were currently functioning...
below grade level, the students were believed to have the intellectual ability to function with their age peers.

The core team for each student's assessment consisted of members of the school district's technology team, including a speech-language pathologist or special education teacher, an occupational therapist, and one of the investigators. All support teams included the student's mother. In addition, M.'s support team included his special education teacher, and both B.'s and D.'s support teams included their occupational therapy aide.

After the students were identified for the TTAP, their parents as well as professionals from their school teams were contacted and informed as to the purpose and procedures of this study. Signed parent consent forms were then obtained.

As called for in the TTAP protocol, a 10-min to 15-min preassessment videotape of each student participating in different activities was made at his school. Selection of preassessment activities was left to the judgment of the investigators. The students were videotaped during activities that involved writing, computing, and another fine motor task. Each core team used the preassessment videotape to plan individualized assessment activities and a tentative assessment agenda. Input from the student's family and school team members was also used to guide the assessment agenda.

After the preliminary assessment activities, the TTAP was administered to each student separately by his respective core team. All assessment activities were videotaped, including the core and support team preparatory meetings. The videotape of B, served as a TTAP training tape for the raters. Interrater reliability was subsequently calculated on the basis of D.'s and M.'s tapes. Each student's assessment lasted approximately 1.5 hours.

Procedures for Rater Training

The raters received 3 full days of training on the use of the TTAP from the instrument's developers. They then received 1 additional day of training, which included a review of TTAP procedures, instruction on how to record data on the Tech Access and Individual Trial Forms, and practice in completing the forms and making recommendations on the basis of B.'s videotaped assessment. The raters worked together to make assistive technology recommendations until 100% agreement was reached within the group. Their final recommendations during the training session were identical to the recommendations made by B.'s core team members after the TTAP assessment.

Data Collection

The raters viewed and scored D.'s and M.'s videotaped assessments on the day after the second TTAP training session. The raters were randomly assigned to one of two groups for the purpose of viewing videotapes. The order in which the groups viewed the tapes was also randomly determined. Before viewing a tape, the raters were told the age of the student being assessed, the reason for the assessment, and the sequence of activities used during the assessment. Each rater also received a separate folder for each student that contained blank Tech Access and Individual Trial Forms. After viewing the videotape, raters independently completed the assessment forms. Discussion among raters was not permitted during the rating sessions.

In accordance with TTAP protocol, raters were free to independently review segments of the videotapes for clarification as needed. They were able to change the data recorded on the Tech Access Form while reviewing a videotape. After all the raters in a group completed the forms, the group viewed the second student's videotape. The same data collection procedure was repeated for the second student.

The investigators collected all materials used by the raters immediately after each rating session. One investigator then conducted individual interviews with each rater to clarify his or her written recommendations, review decision-making processes, and identify subjective reactions to the TTAP. Identical procedures and questions were used during each interview.

Data Analysis

Intraclass correlation coefficients (ICCs) were calculated on the basis of the dichotomous data from the raters' Tech Access Forms (Barthko & Carpenter, 1976). A 1 was recorded when the student demonstrated the TTAP-prompted characteristic or behavior, and a 0 was recorded when the student did not demonstrate the characteristic or behavior. Data were analyzed from the four sections of this form that were applicable to the students: Keyboard Input, Touch Tablet Input, Recommendations for Touch Tablet Input, and Recommendations for Keyboard Input.

Six ICCs were calculated on the basis of which sections of the instrument were applicable to the student. From D.'s data, ICCs for all four sections were calculated. From M.'s data, ICCs were calculated for Keyboard Input and Recommendations for Keyboard Input.

In addition to the ICC calculations, qualitative data were collected from each rater's informally written narrative on the assessment forms and during subsequent interviews. The transcribed interviews and raters' written narratives were analyzed to identify recurring themes (Guba & Lincoln, 1989; LeCompte & Preissle, 1993).

Results

ICC results indicated low consistency among the five
raters on three of the four sections of the Tech Access Form for both students, suggesting poor interrater reliability (see Table 1). Significant results were found only on the Keyboard Input section for both D.’s and M.’s assessments. All other sections were found to be nonsignificant, indicating that rater responses were highly inconsistent.

The ICCs calculated during this study ranged from -.25 to .37. The possible range of scores for an ICC is between -1.0 and +1.0, with a value of 1.0 indicating perfect agreement among raters and 0 indicating chance agreement. A value of less than 0 is difficult to interpret but is not viewed favorably (Cohen, 1960). Additionally, the ICC only corrects for chance agreement but does not identify error source, which contributes to the difficulty in interpreting the ICC.

Regarding specific assistive technology solutions for D., all five raters recommended a standard keyboard, and four recommended enlarged letter stickers on the standard keyboard keys. One rater recommended the use of a word-processing program with auditory feedback, whereas three recommended further evaluation of auditory feedback. Three raters recommended word prediction software for further evaluation. For M., four raters recommended a small, alternative keyboard, and the fifth recommended further evaluation of this small keyboard option. Two raters recommended the use of a standard keyboard; two recommended enlarged letter stickers for the keyboard keys; and two recommended the use of a wrist rest.

Discussion

A goal of this study was to extend the work of the original TIAP developers by testing the assessment’s reliability with older students who had learning disabilities. Although the initial interrater reliability for recommendations generated from the TTAP was reported to be high (98% agreement among 10 raters) (Hutinger, 1993), a closer subsequent examination revealed findings that paralleled our study. High interrater agreement was present only for the final technology recommendations, not for each test item or section of the TTAP (P. L. Hutinger, personal communication, April 22, 1994). In fact, an item-by-item analysis of Hutinger’s data revealed that interrater agreement ranged from 10% to 100%, indicating inconsistency between raters and bringing into question the reliability and utility of the TTAP as an assessment tool even for young students with considerable disabilities.

Two major questions arise on the basis of these findings: Why were the final assistive technology recommendations so consistent when the data from which the recommendations were supposed to be made were so inconsistent? How and to what extent do the actual TTAP assessment items relate to a recommendation for assistive technology, and is other information being used by the raters to guide recommendations? To address these questions, we will discuss the statistics used in our study, characteristics of the raters, characteristics of the students, and possible alternative explanations for the findings.

Statistics Used

Reliability was calculated with an ICC, which Bartko and Carpenter (1976) recommended when five or more raters are used and the data are dichotomous. This analysis presumes that independent raters rate independent subjects. For this study, TTAP items (not students) served as subjects, and some inter-item dependency may have been present. To address this concern, we examined the raw data to evaluate the presence of regular rating patterns across raters. We found that patterns were not present among raters, suggesting that each item affected each rater differently and was, therefore, somewhat independent. Similarly, there may be a question about our use of an ICC versus a generalized kappa statistic. Bartko and Carpenter stated that “a generalized kappa formula is also suitable for dichotomous data and more than two raters. The resulting reliability coefficient is equal to the ICC, but the calculations are more complex” (p. 311).

Although the ICC method was used, findings were very similar to the initial TTAP interrater reliability study findings that were based on percent agreement (Hutinger, 1993; P. L. Hutinger, personal communication, April 22, 1994). This indicates that the different statistical methods used to evaluate interrater reliability in the two studies did not seem to affect the overall findings. Therefore, attention must focus on weaknesses within the instrument itself.

Our finding that the Keyboard Input section of the Tech Access Form yielded the highest interrater reliability indicates that raters used this section more consistently than any other. This suggests that the Keyboard Input section may be adequately designed and easy to use. It is

| Table 1: Intraclass Correlation Coefficients for Four Sections of the Tech Access Form From Student D.’s and Student M.’s TTAP Assessments |

<table>
<thead>
<tr>
<th>Section</th>
<th>ICC</th>
<th>Number of Items Scored</th>
<th>Number of Items Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touch Tablet Input Recommendations for</td>
<td>-.25</td>
<td>8</td>
<td>NA</td>
</tr>
<tr>
<td>Keyboard Input Recommendations for</td>
<td>-.33*</td>
<td>20</td>
<td>.37* 19</td>
</tr>
<tr>
<td>Keyboard Input Recommendations for</td>
<td>-.12</td>
<td>12</td>
<td>-.02 12</td>
</tr>
</tbody>
</table>

*Significant at the .05 level.
important to note that the ICCs for the Keyboard Input sections for both students, although significant, were relatively low.

Although no universally accepted levels have been recognized for evaluating correlation coefficients for the purpose of estimating reliability, Currier's (1990) guideline for qualifying the numerical value of the ICC is .90 to .99 for high reliability, .80 to .89 for good reliability, .70 to .79 for fair reliability, and .69 and below for poor reliability. Had we used Currier's guideline instead of a test for significance, all sections of the Tech Access Form would have shown poor interrater reliability.

**Rater Characteristics**

In addition to an examination of the statistics, it is important to consider the raters' educational and professional characteristics. Because all five raters were specifically trained in the use of the TTAP instrument and had extensive experience in providing assistive technology services to students with disabilities, they may not represent the population of professionals that typically use this assessment. In fact, professionals with far less or different training and experience are likely to use the TTAP. Additionally, the raters' awareness that the data they generated would be analyzed for an interrater reliability study may have influenced the results in a positive direction (Currier, 1990). The low, but significant correlation obtained on the Keyboard Input section of the Tech Access Form suggests that if the instrument had been used by professionals with less training and experience, even lower interrater agreement may have been obtained.

**Student Characteristics**

The students assessed in this study were diagnostically and developmentally different from the students for whom the assessment was originally designed (i.e., students ages birth to 8 years with moderate to severe disabilities, including sensory, motor, or cognitive limitations). Although the TTAP developers suggested that the instrument may have broader application in schools, one could argue that the Tech Access Form was more detailed and specific than was necessary for the students in this study. The raters may have become lost in the detail of the form and confused about how to record on sections that were not relevant to the students being assessed. The low interrater reliability reported for this study, therefore, may be partially due to an instrument that did not adequately match the characteristics of the participating students who had learning disabilities.

**Possible Alternative Explanations**

The most obvious explanation for the discrepancy between the inconsistent ratings on individual TTAP items and the highly consistent final recommendations is that other unmeasured variables were operating to guide rater decisions. The raters' prior knowledge of technology and similar experiences may explain their similar recommendations. This explanation suggests that the recommendation section of the Tech Access Form is independent of the assessment items on which recommendations are supposed to be based. Content validity of the instrument is therefore called into question. Future studies may consider extensive debriefing of raters to identify assessment items that they believe to be particularly important and to identify factors outside the instrument that strongly influence their decision making.

**Conclusion**

Findings from this study suggest that there are serious problems with the TTAP. Additional work related to assessment items, format of the tool, and scoring criteria are needed. This work must be followed by rigorous reliability and validity testing. Only after the TTAP has demonstrated statistical reliability and validity can it be used with confidence.

**Acknowledgments**

We thank the staff members of the Assistive Technology Resource Center at Colorado State University for their extensive support and the students and families for their participation in the assessment sessions. This research was supported, in part, by a grant from the Office of Special Education and Rehabilitative Services, U.S. Department of Education (grant #H1029F30064). This article is based on a thesis submitted to the Academic Faculty of Colorado State University in partial fulfillment of the requirements for the degree of Master of Science in Occupational Therapy.

**References**


