The Relationship Between the Evaluation Tool of Children’s Handwriting and Teachers’ Perceptions of Handwriting Legibility

Pimjai Sudsawad, Catherine A. Trombly, Anne Henderson, Linda Tickle-Degnen

Key Words: education • measurement scales • validity of tests

Meaningful treatment outcome indicators are not absolute measures but, rather, are related to social norms and expectations (Keith, 1995). A measurement of task performance in a contrived setting using a specific measurement tool may not be sufficiently meaningful to represent performance of that same task in an actual context where social norms and expectations will likely play a role in judging the level of performance. As therapists and researchers, we need to understand the connection between what is being measured in therapy and the actual performance in real-life contexts, such as home or school settings. We use standardized tests as a measure of therapy or research outcome because of their objectivity and concrete nature when compared with subjective evaluations, such as opinions and judgment. However, validating the meaningfulness of the scores from such tests in relation to day-to-day ability in real-life contexts is important. Ecological validity—the functional and predictive relationship between performance on a test and behaviors in real-world setting (Sbordone, 1997)—has not been widely investigated empirically in occupational therapy research. Establishing ecological validity of tests may help therapists and researchers to interpret more accurately and use the scores with confidence as they reflect on the outcome of interest in its social context.

Children’s handwriting performance in school is of par-
The American Journal of Occupational Therapy

The research question of this study, therefore, is: Does a relationship exist between the ETCH scores and the level of handwriting legibility in the classroom setting as identified by teachers? To establish ecological validity, the ETCH needs to predict handwriting behaviors in the classroom setting as ranked by teachers.

Because of its social relevance, teacher’s judgment of handwriting legibility during regular classroom writing tasks was used as a benchmark for ecological validity of the ETCH. The teacher is usually the one who refers children with handwriting difficulties for occupational therapy services and determines whether the treatment outcome is satisfactory within the classroom setting.

Method

Participants

Data used for this study were collected from 45 first-grade students who were identified by their teachers as having handwriting difficulties. These data were collected as part of a larger study investigating the effectiveness of a treat-
ment method to improve handwriting. The children's characteristics in the larger study were as follows: (a) identified by their teachers as having below-average handwriting compared with classmates, (b) had normal muscle tone and full passive range of motion of both upper extremities, (c) had kinesthetic deficits as determined by scoring on the Kinesthetic Sensitivity Test (Laszlo & Bairstow, 1985) at or below the 25th percentile, (d) had appropriate level of attention span during classroom activities, and (e) were not on medication intended to improve attention.

The children were recruited from 24 elementary schools within 2 school districts in the Greater Boston area. Their ages ranged from 6 years 2 months to 7 years 11 months (\( M = 6\) years 11 months, \( SD = 5\) months). The median age was 7 years, whereas the mode age was 7 years 1 month. Thirty boys and 15 girls comprised the sample.

**Instruments**

*The ETCH.* Developed by Amundson (1995), the ETCH is a standardized handwriting assessment designed for use with 6-year-olds to 12-year-olds. The test assesses both manuscript and cursive handwriting. Only the assessment for manuscript handwriting was used in this study because first graders only use manuscript writing during their educational activities. The ETCH was designed to measure both legibility and speed of handwriting with six evaluation tasks that resemble tasks required during classroom participation. The six manuscript writing tasks are as follows:

1. **Writing the alphabet from memory,** where the child is asked to write both lowercase and uppercase letters in manuscript from memory
2. **Writing numerals from memory,** where the child is asked to write the numerals 1 through 12 from memory
3. **Near-point copying,** where the child is asked to copy a five-word sentence from a nearby manuscript model
4. **Far-point copying,** where the child is asked to copy one sentence from a distant manuscript model
5. **Dictation,** where the child is asked to write two five-letter nonsense words, and one five-number zip code dictated letter by letter (or number by number)
6. **Sentence composition,** where the child is asked to compose and write a sentence that contains at least five words.

The total scores of three aspects—letter legibility, word legibility, and numeral legibility—were calculated for each child. These scores came from all tasks combined and were transformed into percentages on the basis of the number of readable letters, words, and numbers against the possible letters, words, and numbers, respectively. The test–retest reliability coefficients for first-grade and second-grade children are .63 for total numeral legibility, .77 for total letter legibility, and .71 for total word legibility (Diekema, Deitz, & Amundson, 1998).

*Teacher's questionnaire.* This measurement tool was developed to obtain teachers' judgments of the children's handwriting performance during classroom activities when compared with other children in the same class. Teachers were asked to answer six multiple-choice questions about the child's overall handwriting legibility compared with his or her peers. One question was aimed at the children's overall handwriting legibility, whereas the other five were aimed at handwriting legibility during specific classroom writing tasks. Those tasks were comparable to the ETCH test tasks. These tasks included copying from a near-point model, copying from a far-point model (e.g., the blackboard), dictation, free writing, and number legibility during mathematics.

Each question contained the same seven choices to describe the child's level of handwriting legibility: much-below average, below average, slightly below average, average, slightly above average, above average, and much-above average. A score was assigned to each choice, ranging from –3 for much-below average to +3 for much-above average. A score of 0 was assigned for average.

**Procedure**

Children whose parents consented to participate were taken from the classroom for ETCH testing. Testing was conducted individually in a quiet and, if possible, secluded place within each school (e.g., unoccupied cafeteria, the library, an unoccupied office). Most children were seated on regular child-sized chairs at child-sized tables. If the chair and table sizes were not exactly suitable, appropriate adjustments were made to ensure proper positioning. Teachers were given the teacher's questionnaire the same day of or the next day after the ETCH testing.

A single scorer blind to the children's identities scored the ETCH. Before scoring, the scorer completed two quizzes provided in the test manual and showed satisfactory competency in scoring on the basis of the competency criterion. The scorer's competency was maintained throughout the scoring period by retaking the same quizzes after every 9 or 10 scorings to minimize deviations from the scoring criteria. The competency was satisfactory in all recalibrations. Scoring was carried out within a period of approximately 2 weeks.

**Results**

Table 1 presents detailed information of the ETCH and teacher's questionnaire scores. The correlation coefficients between the teacher's questionnaire general legibility scores (an item in the questionnaire that indicated the teacher's perception of the child's overall legibility) and the ETCH total word legibility (\( r = -.03, p = .81 \)), total letter legibility (\( r = .09, p = .54 \)), and total numeral legibility (\( r = -.12, p = .41 \)) were low and not significant.
Correlation analyses were also performed to examine the relationship between each ETCH subtask score and its counterpart of the children's handwriting legibility in the same task as judged by teachers. No significant relationships were found (see Table 2).

**Discussion**

Almost no relationship was found between the standardized handwriting measurement (ETCH) and the teachers' perceptions of the children's level of handwriting legibility in the classroom. The lack of agreement between the ETCH and the teachers' perceptions was observed both in general and in specific writing tasks. On closer examination, the ETCH high and low scores were scattered within each level of handwriting difficulty as identified by the teachers (slightly below average, below average, much below average). Some children whom teachers identified as having much below average handwriting legibility in the classroom received moderate to high scores on the ETCH. On the other hand, some children whom teachers indicated as only having slightly below average handwriting legibility did poorly on the ETCH. The ETCH legibility scores varied widely from 0% to 100%, although all the children were identified as having problems with handwriting legibility.

Several interpretations of these findings are possible. First, the lack of significant relationship between the ETCH score and the teacher's questionnaire score may indicate that the performance identified by the ETCH does not correspond to the handwriting skills as judged by the teachers. The children who had lower scores on the ETCH were seldom the children whom the teachers judged to have a lower level of handwriting legibility in the classroom setting. Graham (1986) asserted that the construct of legibility has been defined by its general characteristics rather than by a unitary characteristic, making the definition imprecise. Graham and Miller (1980) indicated that legibility is a composite of simpler elements that are interrelated. Those elements include neatness, uniformity, words and letter proportions, alignment, and formation; consistent slant; and line quality. Graham stated that the meanings of these descriptions of handwriting legibility are not adequately established, and it is not clear what degree of each factor or combination of factors constitutes legible handwriting. As Graham concluded, the ambiguity of these criterion variables can be a source of variability in handwriting evaluation.

According to the ETCH scoring criteria, as long as the letters, words, or numbers themselves are readable, differences between two handwriting samples in these elements will not be reflected in the ETCH scores. On the contrary, it is conceivable that teachers will notice these qualitative differences and judge the samples to be different. The ETCH's inability to reflect differences in writing quality may be one of the reasons for the disagreement between the ETCH and teacher's questionnaire scores as seen in this study.

The findings seem to indicate that the percentages of handwriting legibility alone, as used for the ETCH scoring procedure, may not be adequate to reflect children's handwriting legibility in the classroom setting. On the basis of this first interpretation, qualitative factors, such as uniformity, word and letter proportions and alignment, consistency of slant, and line quality, may need to be incorporated into the ETCH's scoring system for it to reflect more closely children's handwriting ability in the classroom as judged by their teachers.

In a previous study, Grace-Frederick (1998) found that the ETCH scores of the children whom teachers judged to have below-average handwriting legibility were significantly

**Table 1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETCH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total word</td>
<td>45</td>
<td>51.1</td>
<td>25.5</td>
<td>0.0</td>
<td>94.7</td>
</tr>
<tr>
<td>Total numeral</td>
<td>45</td>
<td>65.6</td>
<td>13.3</td>
<td>38.9</td>
<td>89.8</td>
</tr>
<tr>
<td>Total numeral</td>
<td>45</td>
<td>72.4</td>
<td>17.7</td>
<td>17.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Number</td>
<td>45</td>
<td>76.3</td>
<td>18.2</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Near-point (word)</td>
<td>45</td>
<td>61.8</td>
<td>34.1</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Near-point (letter)</td>
<td>45</td>
<td>72.5</td>
<td>18.9</td>
<td>27.7</td>
<td>94.4</td>
</tr>
<tr>
<td>Far-point (word)</td>
<td>45</td>
<td>45.8</td>
<td>34.3</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Far-point (letter)</td>
<td>45</td>
<td>66.3</td>
<td>17.4</td>
<td>5.5</td>
<td>94.4</td>
</tr>
<tr>
<td>Dictation (unit)</td>
<td>45</td>
<td>42.2</td>
<td>31.3</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Dictation (letter)</td>
<td>45</td>
<td>61.6</td>
<td>16.0</td>
<td>26.6</td>
<td>93.3</td>
</tr>
</tbody>
</table>

*Note. ETCH = Evaluation Tool of Children's Handwriting.*

*Two children could not form a sentence for this task; therefore scoring was omitted and total scores adjusted.*
different from those of the children whom the teachers judged to have above-average handwriting legibility, indicating discriminant validity of the ETCH. In Grace-Fredericks' study, children from entire classrooms were included as participants. Therefore, handwriting legibility ranged from much-below average to much-above average. In the present study, all children were below average in their handwriting legibility in their teachers' opinions. Although scores on the ETCH can distinguish the children whose handwriting is below average from those whose handwriting is above average, it may not distinguish the more severe from the less severe.

The second interpretation of these findings is that teachers may have judged severity of handwriting problems on the basis of other components in addition to handwriting legibility. Some of the possible factors that may contribute to these judgments include the child's attention to writing details and the child's attitude toward writing. Although these factors might have affected the teachers' judgment of the children's handwriting quality, they would not have influenced the administration or scoring of the ETCH because only the handwriting product was examined. This speculation is based on comments from the teachers that how much the children paid attention to the writing task and whether they liked to write had an impact on how they wrote. This speculation, however, is not substantiated by empirical data. One might infer on the basis of this interpretation that a teacher's perception is not a valid indication of a child's severity of handwriting problems because of other factors, such as those indicated previously, that can influence that perception. Nevertheless, teacher's judgments are important in the social context. Whether the child will be referred for handwriting evaluation and treatment is almost solely based on the teacher's perception of the child's level of performance.

Although direct evidence is not available to support the notion that teachers' judgments of handwriting legibility may be influenced by other factors, it is well established that many factors can significantly influence teachers' judgments when they evaluate student performances. Briggs (1970, 1980) and Markham (1976) found that the format of an essay presentation significantly affected how the essay would be graded. Teachers tended to give a better grade to the essay that was rendered in better handwriting than the same essay presented less attractively. Chase (1986) found complex interactions among different factors such as a student's race and gender, penmanship, and a teacher's level of expectation. These interactions differentially affected teachers' essay scoring even when the essays were identical in both content and presentation. For example, African-American students with good penmanship whose papers were read under high expectations were given relatively high scores compared with Caucasian students with good penmanship and the same high expectations. The relationship, however, was reversed for students for whom readers had low expectations. African-American students with good penmanship whose papers were read under low expectations received relatively low scores compared with Caucasian students with good penmanship, whose papers were read under the same condition. The researcher concluded that variables such as these in multiple combinations differentially influence the reader. Possibly, the teachers in the present study were influenced by factors other than handwriting when assessing level of handwriting legibility.

A third possible interpretation is that the ETCH score represents a one-time observation of handwriting legibility, whereas the teacher's questionnaire score represents a general observation of the skills over time. Handwriting performance on a single occasion and in an artificial testing environment may not be an accurate reflection of the student's ability. Health factors, motivation, personal problems, or distractions (Graham, 1986) may affect how the student writes on a particular day, resulting in handwriting performance that is different from his or her usual or average level of handwriting legibility as determined by the teacher.

This study needs replication in order to verify the results. Further investigation should be done for the total range of children's handwriting legibility from much-below average to much-above average as judged by teachers so that a complete picture of the ecological validity can be obtained. Different types of teacher ratings of handwriting performance may be used as benchmarks for ecological validity. Additionally, further investigation needs to determine whether the qualitative factors (i.e., uniformity, word and letter proportions and alignment, consistency of slant, line quality) actually contribute to teachers' judgments of handwriting legibility in the classroom setting. Another handwriting measurement tool that assesses these factors may be studied and compared with the ETCH for its relationship with teachers' judgments. The results may help to direct future instrument development that is appropriate for predicting or representing actual performance of handwriting in the classroom setting.

If objective, standardized tools are to identify children's levels of performance or capture meaningful changes in performance after treatment, they need to provide results consistent with performance in the social context. Otherwise, "improvement" as identified by an objective, standardized tool measuring either therapeutic outcome or research outcome may not be valid and may mislead and generate inaccurate information. More studies should be conducted to investigate the relationship between other standardized test scores and performance in a natural setting so that the meaningfulness and usefulness of those test scores may be established.

Acknowledgments

We thank the administrators, staff, and students of the Malden and Medford school districts of the Greater Boston area for their generous help, accommodations, and commitment during the study. This study was partially funded by the Dudley Allen Sargent Research Fund of Sargent College of Health and Rehabilitation Sciences, Boston University, and the National Center for Medical Rehabilitation Research.
of the National Institutes of Health (National Institute of Child Health and Development), grant #T32HD07462. This study is based on the first author's doctoral dissertation.

References


