Reliability and Validity of the Evaluation Tool of Children’s Handwriting–Cursive (ETCH–C) Using the General Scoring Criteria

Sharon Duff, Traci-Anne Goyen

KEY WORDS
- education
- educational measurement
- handwriting
- psychomotor performance
- reproducibility of results

OBJECTIVES. To determine the reliability and aspects of validity of the Evaluation Tool of Children’s Handwriting–Cursive (ETCH–C; Amundson, 1995), using the general scoring criteria, when assessing children who use alternative writing scripts.

METHODOLOGY. Children in Years 5 and 6 with handwriting problems and a group of matched control participants from their respective classrooms were assessed with the ETCH–C twice, 4 weeks apart.

RESULTS. Total Letter scores were most reliable; more variability should be expected for Total Word scores. Total Numerical scores showed unacceptable reliability levels and are not recommended. We found good discriminant validity for Letter and Word scores and established cutoff scores to distinguish children with and without handwriting dysfunction (Total Letter <90%, Total Word <85%).

CONCLUSION. The ETCH–C, using the general scoring criteria, is a reliable and valid test of handwriting for children using alternative scripts.


Students with handwriting dysfunction are frequently referred to occupational therapy and form a considerable proportion of the community or school-based occupational therapist’s caseload. Performance on written tasks at school, including assignments and examinations, can be influenced by a child’s poor legibility and speed (Graham, Weintraub, & Berninger, 2001; Tseng & Cermak, 1993).

Assessment forms an integral part of management with this clinical population. Relatively few standardized instruments are available for clinicians to specifically evaluate handwriting performance (Amundson, 1995; Ziviani & Elkins, 1984). Those available are rarely used by occupational therapists in clinical practice (Feder, Majnemer, & Synnes, 2000). Rather, tests that assess underlying abilities, including the Developmental Test of Visual Motor Integration (Beery, 1997), the Bruininks–Oseretsky Test of Motor Proficiency (Bruininks, 1978), and the Test of Visual Perceptual Skills (Gardner, 1996), are reported to be commonly used to assess children with handwriting dysfunction (Feder et al., 2000).

Reliable and valid standardized assessments of handwriting provide an objective measure of actual handwriting performance. In assessing handwriting, these assessments are preferable to tests of underlying abilities (Feder & Majnemer, 2003; Goyen & Duff, 2005). Standardized assessment of handwriting allows comparison between peers and between pre- and posttreatment scores; they can determine eligibility for services and can be used in research. In their review of handwriting...
assessments, Feder and Majnemer (2003) recommended that therapists use a comprehensive approach to handwriting evaluation that includes a range of handwriting tasks necessary for day-to-day functioning in the class. They concluded that the assessment tools available are not widely used and that validation studies for these tools are lacking.

The type of handwriting scripts that are taught by schools, particularly cursive, can vary across countries and states and even within districts. As a result, handwriting assessments have used different writing scripts, which may limit the utility of handwriting assessments that occupational therapists can select for clinical and research purposes in their own contexts. The Evaluation Tool of Children’s Handwriting (ETCH; Amundson, 1995) is one assessment that includes writing models and scoring criteria that could potentially accommodate differing scripts. The ETCH is used to examine legibility across a variety of functional written communication tasks commonly performed in the classroom, including writing from memory, copying from a model, and self-generated writing. Instead of examining detailed and more specific scoring criteria thought to contribute to legibility—such as letter formation, adherence to lines, and spacing of words and letters—the ETCH incorporates global scoring criteria that are based on overall readability of the writing (Rosenblum, Weiss, & Parush, 2003). For instance, a word is considered illegible if it is not quickly, easily, and correctly read as the intended word; is confused for another word; or contains extraneous forms.

The ETCH offers scoring criteria to assess children on both manuscript (ETCH–M) and cursive (ETCH–C) scripts. Scoring encompasses both general scoring criteria for letters, words, and numbers and specific criteria with examples to assist the scorer. Although the ETCH is based on the D’Nealian script, the general scoring criteria can be used with children who use alternate writing styles. The author of the ETCH, Amundson (1995), has commented that children who were unfamiliar with D’Nealian script were not confused by the model script used in the ETCH. Dennis and Swinth (2001) used the general letter and word legibility scoring criteria of both the ETCH–M and ETCH–C to investigate the association between pencil grasp and length of writing task. Their findings suggested that the ETCH general scoring criteria can be successfully applied to scoring writing tasks other than those included in the ETCH. This assumption, however, requires further investigation.

Limited published information exists regarding the ETCH’s psychometric properties, particularly those of the ETCH–C. Moreover, reliability and validity of the ETCH–C’s general scoring criteria when used with alternate writing scripts has not been explored. This information is needed before the ETCH can be considered a useful tool in future research using the general scoring criteria or for clinicians who assess children who use different writing scripts.

Our purpose in this study was to determine the reliability and aspects of validity of the ETCH–C using the general scoring criteria. We designed the study to examine the following in relation to the ETCH–C: intrarater reliability, interrater reliability, test–retest reliability, discriminant validity, concurrent validity with the Test of Legible Handwriting (TOLH; Larsen & Hammill, 1989), and relationship with teacher’s ratings of handwriting.

Method

Participants

Participants were children in Years 5 and 6 (6th and 7th year of formal schooling) attending mainstream public primary schools in New South Wales, Australia. Approval from the Children’s Hospital Ethics Committee was granted. Permission to conduct the study in schools was obtained from the Department of Education and Training and the Catholic Education Offices.

First, we randomly selected 10 public primary schools from within a 20-km radius of Westmead Hospital and obtained their consent to participate. We asked Year 5 and Year 6 teachers to select two groups of participants. One group of students—case students—had handwriting difficulties and were selected by teachers as having difficulty with handwriting legibility or slowness when writing that interfered with their ability to perform in class. Teachers then selected matched control students from the same class who did not have handwriting problems, were of the same gender, and had the closest birth dates to the participants. Written informed consent to participate in the research was obtained by the teacher from the parents of all students. We permitted a maximum of 2 case participants (and 2 control participants) from any one class to ensure that no particular class was overrepresented. Students identified as having a disability, having repeated a class year, having epilepsy, or having been born prematurely were excluded. Case and control participants were therefore matched for major confounding variables deemed to influence handwriting performance, including gender (Ziviani & Watson-Will, 1998), age (Ziviani, 1995), presence of disability, school class (Graham et al., 2001), and type and amount of handwriting instruction (Edwards, 2003; Graham, Berninger, & Weintraub, 1998).

Instruments

Evaluation Tool of Children’s Handwriting. The ETCH (Amundson, 1995) evaluates the legibility and speed of handwriting of children in Years 2 through 6 and has been
used to evaluate the effectiveness of treatment (Case-Smith, 2002; Sudsawad, Trombly, Henderson, & Tickle-Degnan, 2002). It is a criterion-referenced, standardized assessment that focuses on the readability of letters, words, and numbers at a glance and out of context. The ETCH tests manuscript (ETCH–M) and cursive (ETCH–C) writing styles. Tasks are similar to those required of students in the classroom, including writing the alphabet in lower- and uppercase letters from memory, writing numbers from memory, copying a nonsense sentence from a near- and a far-point distance, writing dictation, and sentence composition. For the purposes of this study, the near-point task sheet and far-point copying wall chart were altered by a graphic artist to accommodate the New South Wales (NSW) Foundation Script that is in general use in New South Wales public schools. A global scoring method is used to assess readability of letters, words, and numerals. The ETCH yields Total Letter, Total Word, and Total Numeral scores, which are expressed as a percentage of total legible letters, words, or numbers. Examples of legible and illegible samples are provided in the test manual to assist the scoring process. In addition, legibility components (e.g., letter formation, size, and spacing) can be analyzed. Pencil-and-paper management tasks related to handwriting can also be evaluated. Scoring tutorials and quizzes are included to increase scoring competency.

The ETCH’s psychometric properties have not been comprehensively established. The test manual (Amundson, 1995) reports intrarater reliability intraclass coefficients (ICCs) for 14 children from a regular class and 15 who were referred to occupational therapy for handwriting problems. ICCs for the cursive version were .89 for Total Letter, .94 for Total Word, and .53 for Total Numeral scores. Many of the ICCs for each of the writing tasks were lower than desired. Consequently, Amundson (1995) recommended that the Total Letter, Total Word, and Total Numeral scores be used rather than the individual task scores. Dennis and Swinth (2001), who examined the association between pencil grasp and length of writing, reported very good agreement for the ETCH’s intrarater reliability using the general scoring criteria (letter legibility percentage of agreement = 96.9%–99.4%; word agreement = 86.7%–100%). However, they did not specifically examine the ETCH–C, and participants used either manuscript or cursive and did not use the ETCH writing tasks. Diekema, Deitz, and Amundson (1998) reported moderate levels of test–retest reliability for the ETCH–M but did not examine the ETCH–C. No studies have investigated the ETCH–C’s intrarater or test–retest reliability.

Koziatek and Powell (2002) examined the ETCH–C’s concurrent validity with teachers’ grades on the basis of personal judgment for 101 typically developing students from Grade 4. Pearson correlation coefficients were moderate for ETCH–C Total Letter ($r = .65$) and Total Word scores ($r = .61$). In addition, their study was designed to identify ETCH–C cutoff scores that discriminated satisfactory and unsatisfactory handwriting. Using receiver operating characteristic (ROC) curves, Koziatek and Powell (2002) determined that on the ETCH–C, Total Letter scores of 81% and Total Word scores of 75% were the best cutoff points to distinguish between satisfactory and unsatisfactory handwriting. This was slightly lower than the 85% cutoff score suggested by Amundson (1995). Construct or criterion-related validity studies are needed (Polena Feder & Majnemer, 2003).

Test of Legible Handwriting. The TOLH (Larsen & Hammill, 1989) is a standardized and norm-referenced assessment of handwriting legibility for children in Grades 2 through 12. Writing samples are elicited from a variety of contexts and compared with graded legibility samples. A legibility quotient is obtained (mean = 100, standard deviation = 15). No accepted gold-standard assessment of handwriting legibility exists for this study’s target age group. Despite this, we selected the TOLH because, like the ETCH, it is designed to assess readability of handwriting using a global approach to scoring.

Teacher’s rating of handwriting. Teachers were asked to give an overall rating of each child’s handwriting on a 5-point scale (very poor, poor, average, good, and very good) to provide an overall impression of the child’s writing in the classroom. Test–retest reliability of this scale has been established as good (weighted $k = .73$; Duff & Goyen, 2001).

Procedure

Participants were assessed with the ETCH–C and TOLH at two points in time, 4 weeks apart. We assessed participants at their school in a small group. The order of tests was randomized to eliminate an order effect, and response sheets were coded to ensure that raters were blind to child, group allocation, school, or first or second test administration. After completion of the data collection phase, one rater, experienced in administering and scoring the ETCH–C and the TOLH, scored all tests. The rater used the general guidelines outlined in the ETCH manual, with three minor changes made to accommodate the New South Wales Foundation Script: (1) “manuscript is written when cursive is requested” was not included; we felt that this would be too difficult to score because the manuscript and cursive styles are very similar; (2) looped descendes were permitted; and (3) for the letter $k$, both the D’Nealian and looped Foundation script styles were permitted.

The same rater scored the tests again after a 4-week period to obtain intrarater scores. A second rater, a novice scorer of the ETCH, then scored each initial test, allowing evaluation of intrarater reliability.
**Data Analysis**

We analyzed data for the ETCH–C using the Total Letter, Total Word, and Total Numeral scores. We calculated ICCs to examine intrarater, interrater, and test–retest reliability. An ICC of 0.9–1.0 was considered very high reliability; 0.7–0.9, high reliability; 0.5–0.7, moderate reliability; and 0.3–0.5, low reliability (Hinkle, Wiersma, & Jurs, 1998).

We undertook further analysis of test–retest reliability using a means-versus-difference plot, as described by Bland and Altman (as cited in Peat, 2002). The x-axis represents the average of the test and retest scores. The y-axis plots the actual difference between the two scores. The means-versus-difference plot allows one to observe the spread of error and determine whether systematic error has occurred.

We examined discriminant validity using receiver operating characteristic (ROC) curves to measure the test’s accuracy in separating the participants into those with handwriting problems (case participants) and those without handwriting problems (control participants). A ROC curve is constructed from the sensitivity and specificity calculations of a test. The ROC curve plots the false positive rate on the x-axis and the 1 – false positive rate on the y-axis. It shows the trade-off between a test's sensitivity and specificity. The coordinates of the curve closest to the upper left side indicate the best cutoff point for the test to identify children with handwriting problems. The area under the curve can be used to measure the test’s accuracy (Peat, 2002). If the area under the ROC curve is close to 1, the test’s accuracy in identifying handwriting difficulty is considered excellent. If the area under the curve is closer to 0.5, the test is considered to have poor discriminant ability. An approximate guide for classifying accuracy is 1.0–0.9 = excellent, 0.9–0.8 = good, 0.8–0.7 = fair, 0.7–0.6 = poor, and 0.6–0.5 = fail (Tape, n.d.).

We determined concurrent validity using Pearson’s correlational analysis to measure the association of the ETCH–C Total Letter scores with the TOLH Legibility Quotient scores. Finally, we examined the relationship between the teacher’s rating of the child’s handwriting and the child’s performance on the ETCH–C. Teachers’ ratings were collapsed into three groups: very poor–poor, average, and good–very good. We then compared the differences in ETCH scores among the three groups using analysis of variance and Tukey post hoc analysis. We set the level of significance at \( p < .05 \).

**Results**

The study participants were 63 children from 10 schools. Of these, 33 were case participants and 30 were control participants. Three control participants were not available on the initial assessment day. There were 46 male students and 17 female students, with 32 from Year 5 and 31 from Year 6. Twenty five were age 10, 26 were age 11, and 12 were age 12. The group consisted of 57 right handers and 6 left handers.

Fifteen participants were not available for the retest assessment, leaving 24 correctly matched case–control pairs for the test–retest and validity analyses.

**Reliability**

Intrarater, interrater, and test–retest reliability coefficients for the ETCH–C are reported in Table 1. Agreement for the Total Numeral scores was moderate to low on all three measures of reliability. Results for the test–retest reliability of the Total Letter and Total Word legibility scores were lower than expected.

We constructed means-versus-difference plots for the Total Letter (Figure 1) and Total Word (Figure 2) scores to determine whether any systematic error occurred and to examine spread of error. Both plots illustrate that no systematic error occurred; however, more variability existed for those participants with lower scores. We did not construct a means-versus-difference plot for the Total Numeral scores because of the very low ICCs obtained. The Total Letter plot (Figure 1) showed a narrow range of scores from approximately 80 to 95, and the actual difference between test and retest scores was as much as ±10 points. The ICC may have been artificially lowered because of the group’s narrow range of scores (20 points), with between 80 and a maximum of 100 percentage points possible. The larger differences occurred for participants with lower scores, suggesting more variability in scores for children with poorer handwriting legibility. The Total Word plot (Figure 2), however, showed a wider range of scores (between 50 and 100), and although many participants had little difference between test and retest scores, the difference was as much as ±20 points for several participants. We also found larger differences for participants who achieved lower word legibility scores.

**Discriminant Validity**

We found a Total Letter score of 92 to be the best cutoff point to discriminate between case and control participants (sensitivity = .88 and specificity = .83, as noted in Table 2).

<table>
<thead>
<tr>
<th>Score</th>
<th>Intrarater (n = 63)</th>
<th>Interrater (n = 63)</th>
<th>Test–Retest (n = 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Letter</td>
<td>.80</td>
<td>.84</td>
<td>.61</td>
</tr>
<tr>
<td>Total Word</td>
<td>.71</td>
<td>.62</td>
<td>.65</td>
</tr>
<tr>
<td>Total Numeral</td>
<td>.55</td>
<td>.57</td>
<td>.24</td>
</tr>
</tbody>
</table>

Table 1. Reliability (Intraclass Correlation Coefficients) of the Evaluation Tool of Children’s Handwriting–Cursive (Modified)

January/February 2010, Volume 64, Number 1
The ROC curve (Figure 3) indicates good discriminant validity (area under the curve = .86; 95% confidence interval [CI] = .75–.98).

We found a Total Word score of 85 to be the best cutoff point to discriminate between case and control participants (sensitivity = .71 and specificity = .75, as noted in Table 3). The ROC curve (Figure 4) indicates good discriminant validity (area under the curve = .85; 95% CI = .47–.96).

We found a Total Numeral score of 95 to be the best cutoff point to discriminate between case and control participants (sensitivity = .42 and specificity = .88, as noted in Table 4). The ROC curve (Figure 5) indicates fair discriminant validity (area under the curve = .76; 95% CI = .63–.90).

**Concurrent Validity**

We found the concurrent validity of the ETCH–C Total Letter score with the TOLH Legibility Quotient to be good ($r = .6, p < .001$), using Pearson correlation coefficients.

**Teacher’s Rating and the ETCH–C**

We received teacher’s ratings for only 46 children and categorized their rating of overall handwriting ability into three groups. Twenty-two children were rated as having very poor–poor handwriting, 8 were rated as having average handwriting, and 16 were rated as having good–very good handwriting; two teacher ratings were not returned. We
found a significant difference between groups on all legibility scores (refer to Table 5). Post hoc analyses indicated that the very poor–poor group scored significantly lower than the other two groups, but the average and good–very good groups did not differ. For the Total Numerals score, the very poor–poor group’s performance was significantly lower than that of the good–very good group.

Discussion

In this study, we investigated whether use of the general scoring criteria for the ETCH–C is a reliable and valid measure of legibility for children who use the New South Wales Foundation Script rather than the D’Nealian script. Findings have implications for clinicians and researchers who assess the handwriting of children who use alternate writing scripts.

Results indicate that Total Letter scores are the most reliable and are preferable to use when diagnosing and evaluating handwriting dysfunction. Although Total Word scores are useful, more variability should be expected. Because only a few words actually contribute to the total percentage of scores on the ETCH, any discrepancy is likely to create a larger difference or error. For instance, if two words are written too close together, then both of these words are considered illegible. In relation to interrater reliability, one rater may consider these words to be too close and therefore give a score of 15 of a possible 17, thus leaving a Total Word score of 88%. If another rater considers these words not to be too close and therefore not illegible, he or she would give a score of 17, giving a Total Word score of 100%. Thus, one difference in scoring word legibility would yield a real difference of 12 points between the raters for this participant.

Table 2. Discriminant Validity Using Total Letter Score

<table>
<thead>
<tr>
<th>Score</th>
<th>Case Participants (With Handwriting Difficulties)</th>
<th>Control Participants (Without Handwriting Difficulties)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letters &lt;92</td>
<td>21</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>Letters ≥92</td>
<td>3</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Note. Sensitivity = .88; specificity = .83; positive predictive value = .84; negative predictive value = .87.

Table 3. Discriminant Validity Using Total Word Score

<table>
<thead>
<tr>
<th>Score</th>
<th>Case Participants (With Handwriting Difficulties)</th>
<th>Control Participants (Without Handwriting Difficulties)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word &lt;85</td>
<td>17</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Word ≥85</td>
<td>7</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Note. Sensitivity = .71; specificity = .75; positive predictive value = .74; negative predictive value = .72.

Figure 3. Receiver operating characteristic (ROC) curve for Total Letter score.
The ETCH–C demonstrated reasonable discriminant and concurrent validity. Using the general scoring criteria, we found the test to have adequate to good ability to discriminate between children with and without handwriting dysfunction. Also, we identified cutoff scores for Total Letter (<90%) and Total Word (<85%) scores to distinguish between children with and without handwriting dysfunction.

The general scoring criteria were able to distinguish between those with and without handwriting dysfunction, lending support for use of the ETCH–C as a diagnostic tool. The test did not discriminate between average and above-average writers, as rated by teachers. Although this test feature may not be needed for clinical populations, it could limit the test’s utility for research purposes. These findings, however, should be validated in a larger sample.

This study’s results suggest that the most appropriate cutoff points to differentiate between children with and without handwriting dysfunction is 90% for Total Letter score and 85% for Total Word score. In their validity study of the ETCH–C, Koziatek and Powell (2002) reported that a cutoff percentage score of 82% for Total Letter score and 75% for Total Word score discriminated between satisfactory and unsatisfactory handwriters. In comparison, our cutoff points are higher. Koziatek and Powell’s (2002) study included only fourth graders who were younger than our participants and had less practice writing with a cursive script. One would anticipate that our participants, who have had more practice writing in cursive, would score higher. Moreover, higher cutoff points could be attributed to use of the general scoring criteria, which may not be as precise a measurement as the criteria specified in the ETCH manual and used in the Koziatek and Powell (2002) study.

The TOLH is another handwriting assessment that uses general scoring criteria and has cutoff points to diagnose handwriting dysfunction. Our results revealed that the ETCH–C has good concurrent validity with the TOLH when using the general scoring criteria. The ETCH–C, however, has advantages over the TOLH. First, the TOLH is no longer in print. Also, younger children frequently produce minimal written work in response to the TOLH stimulus

---

Table 4. Discriminant Validity Using Total Numeral Score

<table>
<thead>
<tr>
<th>Score</th>
<th>Case Participants (With Handwriting Difficulties)</th>
<th>Control Participants (Without Handwriting Difficulties)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeral &lt;95</td>
<td>10</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Numeral ≥95</td>
<td>14</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>24</td>
<td>48</td>
</tr>
</tbody>
</table>

*Note. Sensitivity = .42; specificity = .88; positive predictive value = .77; negative predictive value = .6.*

---
picture, which makes scoring difficult. The ETCH–C is designed to examine common handwriting errors, which can be helpful in treatment planning, but the TOLH contains no scope for this. Finally, in addition to examining errors that can influence legibility, the ETCH gives good information regarding functional written communication in the class, which is useful in planning intervention. For instance, therapists can determine whether the child has problems with all written tasks or only with self-generated writing. Error patterns, as identified by the ETCH, can be targeted for intervention.

Figure 5. Receiver operating characteristic (ROC) curve for Total Numeral score.

Table 5. Teacher’s Rating of Handwriting and Evaluation Tool of Children’s Handwriting—Cursive (Modified)

<table>
<thead>
<tr>
<th>Score</th>
<th>Very Poor-Poor (n = 22)</th>
<th>Average (n = 8)</th>
<th>Good-Very Good (n = 16)</th>
<th>p†</th>
<th>F (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letters (M ± SD)</td>
<td>85.2 ± 5.8b</td>
<td>91.6 ± 7.1</td>
<td>94.5 ± 5.2</td>
<td>&lt;.01</td>
<td>12.5 (2, 43)</td>
</tr>
<tr>
<td>Words (M ± SD)</td>
<td>68.7 ± 19.0b</td>
<td>88.8 ± 9.4</td>
<td>90.4 ± 7.2</td>
<td>&lt;.01</td>
<td>11.2 (2, 43)</td>
</tr>
<tr>
<td>Numbers (M ± SD)</td>
<td>92.0 ± 5.7c</td>
<td>95.5 ± 3.3</td>
<td>98.0 ± 2.5</td>
<td>&lt;.01</td>
<td>8.6 (2, 43)</td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation.
†Analysis of variance with Tukey’s post hoc test.
‡Significantly lower than other groups.
§Significantly lower than good–very good group.

Numerical legibility is crucial for mathematics and can mean the difference between correct or incorrect answers. Total Numeral scores on the ETCH, however, showed unacceptable levels of reliability and validity and should be used with caution.

In relation to using the ETCH–C’s general scoring criteria to evaluate treatment outcomes, our results are not clear. Test–retest reliability was lower than desired but within the moderate range and useful for clinical purposes. This test characteristic is important to determine whether change in scores over time is more likely the result of treatment provided than of test error or instability. A study by Diekema et al. (1998) found moderate test–retest reliability on the ETCH–M (Letter ICC = .77, Word ICC = .71, Number ICC = .63) in a group of children with handwriting dysfunction. In response to the Diekema et al. (1998) study, Schneck (1998) speculated that test–retest reliability may have been lower than expected because performance is known to be more inconsistent among children with difficulties. On examining our test–retest data closely, we determined that scores may have been artificially lowered because of their narrow spread. When using the general scoring criteria, we found a test–retest error of up to 10 points. This finding could be validated in a subsequent study using a shorter retest period.
The study’s major limitation is the selection of participants based on teacher judgment, which assumes that teacher perception of handwriting is accurate and consistent. Concerns have been raised about the reliability of teacher’s judgments of handwriting (Daniel & Froude, 1998; Sudsawad, Trombly, Henderson, & Tickle-Degnen, 2001). Despite these concerns, other studies have relied on teacher judgment to identify participants with handwriting dysfunction (Diekema et al., 1998; Sudsawad et al., 2001; Wallen & Mackay, 1999). Others have found experienced teachers to be good judges of handwriting legibility (Koziatek & Powell, 2002; Tseng & Murray, 1994).

Conclusion

We investigated the ETCH–C’s reliability and aspects of validity when the general scoring criteria are used. Results indicate that Total Letter scores are most reliable and that although Total Word scores are useful, more variability should be expected. Total Numeral scores showed unacceptable reliability levels, and we would not recommend their use. The test–retest reliability coefficients were, however, lower than desired, similar to many other pediatric tests. We found the ETCH–C to have adequate to good ability to discriminate between children with and without handwriting dysfunction, with established cutoff scores for Total Letter of <90% and for Total Word of <85%. These cutoff scores could be verified in future research using different age groups.

This study confirms the use of the ETCH–C’s general scoring criteria, particularly that for the Total Letter score because it is most reliable. This confirmation is helpful to clinicians when they are assessing children who use an alternate writing script in terms of identifying children with handwriting dysfunction and in treatment planning. The ETCH–C would be a useful research tool to quantify differences between groups and to distinguish children with handwriting dysfunction using the cutoff scores.

Our results reflect the subjective nature of scoring handwriting legibility, particularly with assessments using a global scoring method. The ETCH has been designed specifically to accommodate a more general process of scoring rather than requiring precise measurements with, for example, a ruler. Scores should always be interpreted as part of a comprehensive evaluation of a child’s handwriting skills.

Acknowledgments

We would like to acknowledge Margaret Wallen for her assistance with the study design and manuscript preparation and Jennifer Peat for her statistical support and advice.

References


