Validating the Use of the Evaluation Tool of Children’s Handwriting–Manuscript to Identify Handwriting Difficulties and Detect Change in School-Age Children

Marie Brossard-Racine, Barbara Mazer, Marilyse Julien, Annette Majnemer

In this study we sought to validate the discriminant ability of the Evaluation Tool of Children’s Handwriting–Manuscript in identifying children in Grades 2–3 with handwriting difficulties and to determine the percentage of change in handwriting scores that is consistently detected by occupational therapists. Thirty-four therapists judged and compared 35 pairs of handwriting samples. Receiver operating characteristic (ROC) analyses were performed to determine (1) the optimal cutoff values for word and letter legibility scores that identify children with handwriting difficulties who should be seen in rehabilitation and (2) the minimal clinically important difference (MCID) in handwriting scores. Cutoff scores of 75.0% for total word legibility and 76.0% for total letter legibility were found to provide excellent levels of accuracy. A difference of 10.0%–12.5% for total word legibility and 6.0%–7.0% for total letter legibility were found as the MCID. Study findings enable therapists to quantitatively support clinical judgment when evaluating handwriting.

The prevalence of handwriting difficulties has been estimated at 10%–30% in typically developing school-age children (Karlsdottir & Stefansson, 2002). Children with delays or impairments in the development of these skills are frequently referred to occupational therapy for evaluation and intervention (Chandler, 1994; Oliver, 1990). Although computers are now commonly used in classrooms, traditional handwriting is still the primary medium of written expression required by teachers in the school environment (Latio, 2009).

Handwriting is a challenging construct to measure. Although many standardized evaluations exist (see Rosenblum, Weiss, & Parush, 2003, for a review), pediatric occupational therapists do not tend to use standardized assessments of handwriting and prefer to use standardized tests to evaluate the underlying components of the task rather than the handwriting performance itself (Crowe, 1989; Feder, Majnemer, & Synnes, 2000; Rodger, 1994). One reason for this practice may be that handwriting is a complex task involving many sensorimotor, cognitive, and psychosocial components that must be comprehensively evaluated before designing a treatment plan (Schneck & Amundson, 2010). In addition, most standardized handwriting evaluations involve only a single task, usually copying or sentence composition, which might not be sufficient to capture the many aspects of a child’s handwriting performance (Feder & Majnemer, 2003).

The Evaluation Tool of Children’s Handwriting (ETCH; Amundson, 1995) is the only Latin alphabet–based standardized measure that assesses...
a range of handwriting tasks similar to those experienced in the classroom. It takes approximately 30 min to administer and is designed to identify and characterize handwriting difficulties in young school-age children. The ETCH is available in both manuscript and cursive versions. Letters, numerals, and words are judged for legibility using a list of specific criteria such as omission, closing, misplacing, reversion, and poor erasure. The percentage of legibility is determined for each task by counting the legible letters, numerals, or words and dividing by the total number of letters, numerals, or words required. The percentages from each task are then averaged to provide a total legibility score for letters, numerals, and words. The manuscript version of the ETCH (ETCH–M) targets children in Grades 1 to 3 and examines legibility through seven different tasks: alphabet writing from memory (upper- and lowercase), numeral writing from memory, near-point copying, far-point copying, dictation of nonwords (nonsense words) and numbers, and composition of a short sentence. Performance time or writing speed is measured in seconds for the alphabet and numeral writing and in letters per minute for the copying and composition tasks. The variety of tasks makes this criterion-referenced measure the most ecological evaluation of handwriting (Feder & Majnemer, 2003).

Content validity was established in three pilot editions of the ETCH (Amundson, 1995). The concurrent validity of the ETCH–M, however, has been questioned. Sudsawad, Trombly, Henderson, and Tickle-Degnen (2001) obtained a low and insignificant level of correlation between ETCH–M scores and teacher judgments and found that the ETCH–M was not sensitive enough to identify the level of difficulty experienced by handwriting physicians whom teachers perceived to be below average. Nevertheless, an unpublished study (Grace-Frederick, 1998) provided evidence that the ETCH–M was able to discriminate between second-grade children who had handwriting performance that was judged by teachers to be either below or above average. Unfortunately, no cutoff value has been published to identify children with below-average performance on the ETCH–M. Moreover, the concurrent validity of the ETCH–M with the perceptions of rehabilitation clinicians working with children who experience handwriting difficulties has not been reported.

Clinicians often use the ETCH–M as a measure of change before and after an intervention. Test–retest reliability indicators for the ETCH–M are $r = .63$ for total numeral legibility, $r = .77$ for total letter legibility, and $r = .71$ for total word legibility. Total scores for letter and word legibility were found to be more consistent than the individual task scores (Diekema, Deitz, & Amundson, 1998). The interrater reliability with intraclass correlation coefficients (ICCs) ranged from .42 to .84 for the individual items on the ETCH–M for children referred to occupational therapy (Amundson, 1995). Specifically, the ICC was .48 for total word legibility and .84 for total letter legibility. To our knowledge, responsiveness of the ETCH–M has never been reported. In general, the ability of a tool to discriminate among clinical groups can be considered preliminary evidence that it may be responsive to change (Revicki, Hays, Cella, & Sloan, 2008). In health-related quality-of-life research, in which self-report questionnaires are used extensively, the construct of a minimal clinically important difference (MCID) was documented. MCID corresponds to the “smallest change in scores that would likely be important from the patient’s or clinician’s perspective” (Revicki et al., 2008, p. 103). This construct can also apply to outcome measures, because they are known to have their share of variance caused by an individual’s personal factors or by the measurement procedure itself (Leemrijse, Meijer, Vermeer, Lambregts, & Adér, 1999). It is thus relevant and important that clinicians and researchers pay more attention to MCID and integrate this construct when measuring change in scores over time. This psychometric construct has not yet been studied for the ETCH.

The purposes of this study were (1) to identify the ETCH–M cutoff values for word and letter legibility that discriminate between children with handwriting difficulties who should be seen in rehabilitation for a comprehensive evaluation or treatment and those who do not require services and (2) to determine the MCID on the ETCH–M letter and word total legibility by determining the percentage of change that corresponds to the change that experienced pediatric occupational therapists consistently visually detect.

**Method**

This cross-sectional study was approved by the Montreal Children’s Hospital–McGill University Health Centre Research Ethics Board. Written consent was obtained before each evaluation.

**Participant Selection**

Participants included a convenience sample of pediatric occupational therapists working in the greater Montreal, Quebec, area. To be eligible for the study, the occupational therapists had to be working with children and had to have at least 1 yr of experience in evaluating and treating handwriting difficulties.
The handwriting samples were obtained from a group of children involved in a cohort study of school-age children with attention deficit hyperactivity disorder (ADHD). This study evaluated handwriting performance at diagnosis and after initiation of stimulant medication (Brossard-Racine, Majnemer, Shevell, Snider, & Bélanger, 2011; Brossard-Racine, Shevell, Snider, Bélanger, & Majnemer, 2012). Children ages 7–9 yr in Grades 2 and 3 who routinely used manuscript handwriting and did not learn cursive writing for a period greater than 2 mo were included in the cohort. Each child was evaluated once before initiation of medication and also completed one or two evaluations following initiation of medication. In total, 26 children produced a total of 35 pairs of handwriting samples (referred to as Sample A [before initiation of medication] and Sample B [after initiation of medication]).

Instruments

ETCH–M. The ETCH–M (Amundson, 1995) was the standardized measure of handwriting used to evaluate the children. An experienced occupational therapist who was unaware of the study hypotheses administered the ETCH–M according to instructions in the examiner’s manual. All samples had been previously scored by an independent corrector who was blind to the children’s identity and to the timing of testing (i.e., whether the sample was obtained before or after medication initiation). The occupational therapist scorer was trained for scoring consistency, as required by the ETCH examiner’s manual. Only the total scores for word legibility and letter legibility were included in the analysis.

Occupational Therapists’ Ratings of Handwriting. The participating occupational therapists completed a questionnaire that asked two questions for each pair of handwriting samples. The first questions referred only to Sample A and aimed to establish consistent cutoff values that would distinguish children with handwriting difficulties who should be seen in rehabilitation for comprehensive evaluation or treatment from those who did not require services. Occupational therapists were asked to answer the following question: “Looking only at Sample A, does this child need handwriting rehabilitation services?” The therapists were directed to view the handwriting sample as though it had been submitted by a teacher or a parent and to respond only on the basis of their visual inspection and clinical experience. The therapists replied with one of the following responses: yes, no, or not sure.

The second question asked occupational therapists to compare the two samples, A and B, using a 5-point Likert scale to indicate the degree of observable change (1 = B is much better than A, 2 = B is somewhat better than A, 3 = B is about the same as A, 4 = B is somewhat worse than A, 5 = B is much worse than A). Change was defined to the therapists as at least a small decrease or increase between two samples, including a large amount of change. Change was independent of the initial performance (e.g., it did not mean that a child no longer had difficulties). The therapist was unaware of the child’s identity or condition, time between the two evaluations, whether any intervention had occurred between the two samples, and legibility scores obtained for each sample. However, they were informed that the samples were all produced by children in Grade 2 or 3, age 7–9 yr, who were attending a regular school. In addition to these two questions, each participant completed a short sociodemographic questionnaire to collect information on their work environment and level of experience.

Procedures

As part of the broader study described earlier (Brossard-Racine et al., 2011, 2012), the ETCH–M was administered by an occupational therapist to the children after obtaining written informed parental consent and each child’s assent. The primary investigator (Marie Brossard-Racine) selected independent samples of handwriting for use in this study. An e-mail with a brief description of the project was sent to the occupational therapy program coordinators in the two major pediatric hospitals and their affiliated rehabilitation center and to the largest pediatric occupational therapy private practice clinic in the province of Quebec. Once an occupational therapist expressed an interest in participating, he or she was contacted by the investigator to further clarify the study objectives and procedures. A time and place of mutual convenience to meet and conduct the study was then arranged. The total time to complete the procedures and the questionnaire was approximately 45 min.

Data Collection

Every occupational therapist met individually with Marie Brossard-Racine to complete the questionnaire. The same sequence of samples was presented to all occupational therapists, who were subsequently directed to complete the two questions for each pair of samples. A research assistant entered all data into a database.

Data Analysis

Descriptive statistics were used to characterize participants’ demographic and clinical attributes. To determine the agreement (intrarater reliability) between the 34 participants on each of the two questions, Shrout–Fleiss
reliability ICCs with random set were calculated (Shrout & Fleiss, 1979). Receiver operating characteristic (ROC) analyses were conducted to evaluate the ability of the ETCH–M total word and letter legibility scores to identify children who required intervention, as determined by the occupational therapists’ perceptions. First, a crude ROC curve was derived by setting cutoff scores on the handwriting scores. The crude area under the curve and its standard error were calculated using the trapezoidal rule described by Hanley and McNeil (1982). In addition, to account for the repeated-measures design of the study, an adjusted ROC curve was derived by setting cutoff probabilities on the predicted probabilities of the generalized linear mixed model with random intercepts for children and occupational therapists. The area under the adjusted curve was estimated using the trapezoidal rule, and the standard error was estimated using a bootstrap procedure (Liu & Li, 2005). For the total word legibility and letter legibility scores, the optimal value was defined as the shortest distance to sensitivity = 1 and 1 − specificity = 0 on the crude ROC curve. The optimal value also corresponded to the value that best discriminated between children with handwriting difficulties who should be seen in rehabilitation and those who did not need services.

Although no consensus exists on the best statistical way to determine MCID, a combination of an anchor-based approach, in which change scores are compared with clinicians’ or clients’ perceptions, and a distribution-based approach, which focuses on the variation of the sample and on the measurement’s statistical properties, is recommended (Guyatt, Osoba, Wu, Wyrich, & Norman, 2002; Hays & Woolley, 2000; Revicki et al., 2008). It is also recommended to determine a range of values for MCID because they are sensitive to variation across participants (Yost & Eton, 2005). In light of these recommendations, we first evaluated the ability of the ETCH–M word and letter legibility scores measured at two time points to identify children whose samples were perceived by experienced occupational therapists as having changed (either negatively or positively) in their handwriting skills using the same methods used for ROC analysis. Of all the possible distribution-based approaches, it has been demonstrated that half a standard deviation at baseline (0.5 standard deviation) is the most reliable statistic to determine MCID (Norman, Sloan, & Wyrwich, 2003, 2004; Revicki et al., 2008); therefore, the results for the anchor-based approach were compared with 0.5 standard deviation for both word and letter total percentage of change. For all analyses, a confidence interval (CI) of 95% was used, and all analyses were performed using SAS (Version 9.0; SAS Institute, Cary, NC) and R (Version 12.0; R Foundation for Statistical Computing, Vienna, Austria).

Results

Participants

The mean age of the 26 children who produced the samples was 8.1 ± 0.8 yr. Mean total word legibility was 70.6% ± 24.9%; the range of values was 20.0%–100.0%; mean total letter legibility was 73.3% ± 13.9%, and the range of values was 32.0%–96.0%. For the 35 pairs of samples that were compared, the range of percentage of change (using the legibility scores of the ETCH) was −22.0% to 45.0% for word legibility and −8.0% to 32.0% for letter legibility; mean change scores were 6.8% ± 14.7% for word legibility and 3.8% ± 9.4% for letter legibility.

Thirty-four occupational therapists (33 women, 1 man) participated in this study. Twenty-one percent of the sample had 1–2 yr of experience working as an occupational therapist with clients who had handwriting difficulties, 32% had 3–5 yr of such experience, and 47% had >6 yr of such experience. Nineteen therapists were primarily based in a private clinic, 7 were in a rehabilitation center, 5 were part of a school board, and 3 were in a hospital setting. For most participants (30 of 34), 76%–100% of their caseload consisted of school-age children. The most common diagnoses or associated conditions in the children treated with handwriting difficulties were learning difficulties (including ADHD), coordination difficulties (including developmental coordination disorder), and autism spectrum disorders.

Discriminative Validity

The ICC obtained for interrater reliability for the question, “Does this child need handwriting rehabilitation services?” was .53. The ROC curves for word legibility and letter legibility are presented in Figures 1 and 2. The crude area under the ROC curve for word legibility was .86 (95% CI = .84, .88), and the adjusted area was .96 (95% CI = .95, .97). We obtained similar results for letter legibility, with a crude area under the curve of .82 (95% CI = .80, .84) and an adjusted area of .96 (95% CI = .95, .97). We ignored the fact that the repeated-measures design of the study produced estimates that were biased toward zero for the area under the ROC curve, as suggested by Liu and Li (2005). The optimal cutoff score in word legibility scores obtained from inspection of the crude ROC curve was 75%, which corresponded to a crude sensitivity of 77% and specificity of 82%. For letter legibility results, an optimal cutoff score...
of 76% simultaneously maximized sensitivity and specificity, with a sensitivity of 78% and a specificity of 66%.

Minimal Clinically Important Difference

**Anchor-Based Approach.** Because the direction of change (worse or better) was not of concern, the absolute change in word and letter scores was used to construct the ROC curves. The crude and adjusted areas under the ROC curves were .60 (95% CI = .58, .62) and .77 (95% CI = .75, .80), respectively, for the absolute change in word legibility and .61 (95% CI = .58, .62) and .77 (95% CI = .75, .80), respectively, for the absolute change in letter legibility. The ROC curves for word legibility and letter legibility are presented in Figures 3 and 4.

The optimal cutoff value for the percentage change in word legibility score was 10.0%, for a sensitivity of 58.0% and a specificity of 58.0%. For the absolute change in letter legibility, we identified 6.0% as the optimal cutoff score, which corresponded to a sensitivity of 62.0% and a specificity of 57.0% to detect the presence of change in handwriting. The consistency of responses across occupational therapists (interrater reliability) was weak (ICC = .19). We conducted subgroup analyses to identify clinical and demographic variables that may have affected the homogeneity of responses; however, we found similar ICCs among subgroups of occupational therapists working in the same environment or with the same number of years of experience.

**Distribution-Based Approach.** We used the simple estimate of half a standard deviation for both criteria and obtained a value of 12.5% as the MCID for total word legibility and 7.0% for total letter legibility. Thus, combining the two approaches, the MCID for total word legibility ranges from 10.0% to 12.5% and for total letter legibility from 6.0% to 7.0%.

**Discussion**

**Discriminative Validity**

The primary objective of this study was to identify cutoff values for total word legibility and letter legibility percentages on the ETCH–M that would discriminate between children whom experienced therapists perceived as requiring rehabilitation services (evaluation or treatment) for handwriting difficulties and those who did not require services. Because of the repetitive design of this study (34
occupational therapists evaluating the same 26 samples), the ROC curve was adjusted to minimize bias (Liu & Li, 2005). The bias appeared to be moderate, however, because the occupational therapists obtained relatively good interrater reliability (ICC = .53). Good agreement between the crude and adjusted areas under the curve was obtained. The high adjusted area under the curve for both total word legibility and total letter legibility implied that these two criteria had excellent accuracy in distinguishing children. According to the pediatric occupational therapists, 75.0% total word legibility and 76.0% total letter legibility discriminated between children with and without handwriting difficulties. These discriminative values are lower than the 85.0% cutoff scores suggested by Admunson (1995) in the ETCH manual. As Admunson explained, however, the ETCH was still a work in progress at the time of publication of the manual, and the suggested values were only a starting point that required validation. Because no other cutoff scores for the ETCH–M have been published, we are unable to compare our results with other findings. Nevertheless, our cutoff values are similar to those reported for the cursive version of the ETCH (75.0% for total word legibility and 82.0% for total letter legibility; Koziatek & Powell, 2002).

Cutoff values are limited; they consider only the handwriting product and do not take into account different factors that can affect performance (e.g., level of experience, time of year, instructional method, environmental factors). Nevertheless, cutoff values provide quantitative evidence to support therapists when performing comprehensive evaluations and justifying the need for intervention (Majnemer & Limperopoulos, 2002). The cutoff values we determined in this study relate specifically to children in Grades 2 and 3. Children’s handwriting usually reaches a plateau by the end of Grade 2, and no significant improvement is expected in the following years (Graham, Weintraub, & Berninger, 2001). Also, none of the children in Grade 2 were tested in the early months of the school year, and we are therefore confident that these cutoff scores can guide clinicians in interpreting legibility for children within this age group. In this study, several occupational therapists verbalized their frustration at being limited to a handwriting sample and not being able to observe the children’s performance. Nevertheless, matching visual perception to scores obtained on a standardized evaluation is an unbiased way to ascertain the discriminative validity of an instrument.

Minimal Clinically Important Difference in Legibility

Identifying the percentage of change in total word legibility and total letter legibility that clinicians consistently detected resulted in lower levels of accuracy. According to the adjusted area under the curve, only fair accuracy to distinguish change was obtained. Considerable differences between the crude and adjusted estimates of the area under the curves were also found. Those differences could be explained by the poor level of interrater reliability found among the occupational therapists (ICC = .19), indicating that the therapists often rated the same sample differently. Subgroup analyses according to workplace and level of experience did not further clarify the inconsistency of ratings among occupational therapists. The therapists who participated in the current study did not score or administer the ETCH themselves; they were asked to judge samples only by looking at written product. We felt this to be the best way to evaluate perceptions of legibility change because the examiners were not potentially biased by any external factors that would have interfered with their perspectives on the child’s handwriting legibility. It was therefore surprising to see the low level of agreement for this question, which contrasted considerably with the level of agreement reached for the question on the children’s handwriting difficulties requiring rehabilitation.

Several explanations are possible for the low level of agreement among occupational therapists in rating clinically meaningful change. First, they were not previously trained for scoring consistency; the likelihood of error was therefore higher. Also, the group of occupational therapists who participated was not homogeneous; it included clinicians with different degrees of experience, from...
different workplaces, who served different clientele even when from the same institution. Another possible explanation is that the question posed to the occupational therapists might have been too broad and nonspecific. It addressed overall change between the two samples and was not specifically focused on change in word or letter legibility.

Several factors are known to interfere with handwriting legibility. Difficulties in letter formation, spacing, and alignment, to name just a few, can affect word and letter legibility differently (Schneck & Amundson, 2010). Change in word and letter legibility between two samples was not always of the same magnitude. For instance, in Sample B, spacing could have been improved, thereby dramatically increasing word legibility; however, letter formation could have been as poor as that of Sample A, and thus letter legibility did not change sufficiently from a clinical perspective. In this type of situation, each occupational therapist had to decide on the overall change, with no directive regarding specific aspects of word or letter legibility. Some therapists could have prioritized change (or lack thereof) in letter legibility over change in word legibility, whereas others may have done the opposite. This lack of guidance may have specifically affected perception of change. The first question on the need for handwriting intervention might have been more stable because it did not take into account the relative severity of difficulties or the child’s level of difficulty with letter or word legibility, because both might appear in the presence of handwriting difficulties.

This weak level of agreement on what constituted a level of change also supported the importance of using a combination of approaches when determining MCID. In our case, we obtained fairly similar MCID values using two methods. It has been proposed that anchor-based values have more weight than distribution-based values (Revicki et al., 2008), and we thus suggest that our findings may be used as a starting point to appreciate change in spite of the fair level of accuracy we reached using the anchor-based method. Nevertheless, more studies on the MCID of the ETCH–M would help confirm the most appropriate range of values.

Finally, the discriminative validity and MCID values found for word legibility must be interpreted with caution because the ETCH–M interrater reliability for total word legibility was somewhat low (ICC = .48). Low interrater reliability may have influenced our ability to find accurate cutoff and MCID values for this specific subscale. Scores found for letter legibility are probably more accurate because the interrater reliability for this domain was higher (ICC = .84).

Implications for Occupational Therapy Practice

For children in Grades 2 and 3, 75.0% total word legibility and 76.0% total letter legibility on the ETCH–M are suggested as the cutoff values to discriminate between children with handwriting legibility difficulties who should be seen in rehabilitation for evaluation and treatment from those who have no such difficulties. With the ETCH–M, clinicians should not consider a change in scores over time as clinically significant if that change corresponds to less than 6.0% for total letter legibility and less than 10.0% for total word legibility. To summarize, the findings of this study have the following implications for occupational therapy practice:

- Preliminary cutoff values of 75.0% for total word legibility and of 76.0% for total letter legibility on the ETCH–M are suggested to discriminate between children in Grades 2 and 3 who present with handwriting difficulties and those who do not.
- These quantitative cutoff values can support clinicians in justifying need for intervention.
- Change in handwriting scores ≤6.0% for total letter legibility and ≤10.0% for total word legibility should not be considered clinically important.

Conclusion

Our results contribute to the documentation of the psychometric properties of the ETCH–M and provide new information on the concurrent (discriminative) validity when comparing total legibility scores with the perceptions of pediatric occupational therapists. Although the cutoff scores demonstrated excellent discriminative abilities, discriminative values do not replace clinical judgment, and clinicians should always perform a comprehensive evaluation of the child’s performance. The MCID levels are essential to measure the real impact of intervention. Further study is needed to determine cutoff values and clinically meaningful change in handwriting legibility scores.

Acknowledgments

Marie Brossard-Racine was supported by a doctoral fellowship from the Fond de Recherche en Santé du Québec (FRSQ) during preparation of this article. She was also supported by an Edith and Richard Strauss Fellowship and a studentship from the Stars/Montreal Children’s Hospital Studentships in early phases of this work. The research laboratory benefited from infrastructure provided by the Montreal Children’s Hospital Research Institute and the Centre de Recherche Interdisciplinaire en Réadaptation, both supported by the FRSQ. Special thanks to all the.
occupational therapists who took the time to participate in this project, to Joey Wakin for research assistance, and to Victoria Surtees for manuscript correction.

References


