Evidence Topic: Handwriting Assessment

January 2008
Primary Reviewer: Wendy Collins, MOTS
EBPX Team Members:
- Catherine Candler, PhD, OTR
- Jamie Sanders, MOTS

Evidence Question: What handwriting assessments are available that have been shown to have consistency and accurately measure what they are testing?

Evaluation Tool of Children’s Handwriting (ETCH)
- Assessment Description
- Evidence Reviewed
- EBPX Findings
- EBPX Strength and Impact Summary

Minnesota Handwriting Test (MHT)
Childrens Handwriting Evaluation Scales (CHES)

Question Background:
What situations inspired this question?

There are a variety of handwriting assessments available to choose from. The focus of this review was on evidence pertaining to the creditability of handwriting assessments. The measure of an assessment lies in its reliability (consistency) and its validity (how well the assessment measures what it is meant to measure). These terms are further defined below:

Parameters of the Search:
It is important to know how thoroughly the literature was searched for research studies concerning the question. If the search was not intensive, important information may be lacking from the review.

Parameters: No adults or rehabilitative studies were considered. The focus was on articles discussing the validity and reliability of the assessments. Only assessments referenced with research were included in the review.

Keywords: handwriting, handwriting assessment, evidence, re-test, reliability, validity, CHES, MHT, Minnesota Handwriting Test, ETCH, Evaluation Tool of Children’s handwriting, TOLH, Test of Legible Handwriting, SCRIPT, Scale of Children’s Readiness in Printing, Handwriting Performance Test, handwriting performance, Handwriting Speed Test, children, school, elementary
Websites, Resources: CINAL, PubMed, Education Research Complete, Medline, SCOPUS

Reliability – measures the consistency, repeatability of an assessment. There are four types of reliability.

1.) Inter-rater – different raters/observers give consistent answers/scores when administering the test.
2.) Test-retest – the stability of the test over a period of time.
3.) Parallel - assess the consistency of the results of two similar tests.
4.) Internal Consistency – assesses the consistency of results across the items within the test.

Validity – determines how well does the assessment test the skill areas it is designed for. There are six types of validity.

1.) Face – The test appears that it should measure what is intended. This is the weakest type of validity.
2.) Content – The content of the test is logical and suitable for the purpose of the test. This is an overview.
3.) Concurrent – the ability of the test to distinguish between groups that it is testing.
4.) Predictive – how well can the test predict outcomes.
5.) Convergent – how similar is the test to others.
6.) Discriminate – how dissimilar is the test to others.

Reliability and validity are intertwined. Assessments need to have both reliability and validity. For further clarification see

http://www.socialresearchmethods.net/kb/reltypes.php
http://www.socialresearchmethods.net/kb/measval.php
http://www.socialresearchmethods.net/kb/relandval.php

The Research Methods Knowledge Base, 2nd Ed

Evaluation Tool of Children’s Handwriting
http://www.theraproducts.com/

ETCH
Assessment Description:
The Evaluation Tool of Children’s Handwriting (ETCH) developed by Susan J. Amundson has a manuscript and cursive version (Asher, 2007). The manuscript version is for grades 1 to 3 and the cursive is for grades 3 to 6. The ETCH is considered to be a holistic test, meaning it is scored based on the scorer’s overall judgment of legibility. It is criterion-referenced meaning it measurers the individual’s performance based on the individual’s skill rather than comparing the performance to an average. The child must be familiar with manuscript for at least 10 to 12 weeks before the test should be administered (Feder & Majnemer, 2003).

The ETCH has several test domains. These are writing 1) the alphabet (upper and lower case), 2) numbers, 3) near-point and 4) far-point copying, 5) dictation, 6) composition, and 7) speed (Feder & Majnemer, 2003). The test usually takes between 15 to 30 minutes to administer and 10 to 20 minutes to score. Each task is scored based on the objective criteria and exemplars (Asher, 2007). The areas assessed are 1) form, 2) spacing, and 3) size (Tomchek & Schneck, 2007). The baseline data is determined by legibility and speed (Asher, 2007).

**ETCH**

**Evidence Table**

Contains appraisals of evidence reviewed.

### Key to Level of Evidence

(Level of evidence may be adjusted downward if study has poor rigor.)

<table>
<thead>
<tr>
<th>Level</th>
<th>Type of Evidence</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systematic Reviews and meta-analyses</td>
<td>Randomized Control Trials (RCT)</td>
<td>Quasi-experimental and Comparative studies</td>
<td>Correlation and Non-experimental studies</td>
<td>Descriptive studies &amp; Expert Opinion articles</td>
<td></td>
</tr>
</tbody>
</table>

### Citation


### Handwriting Assessments

*ETCH – M*

### Description of Evidence/Type of Study

Test re-test reliability is the consistency of results over time.

### Level of Evidence

3

### Description of Population

18 first and 13 second graders with handwriting dysfunction. Total population = 31 Consent was obtained.

### Description of Intervention

ETCH-M was administered to each student on two separate occasions with a 7 day interval. The same person administered and scored both tests. The administrator was blind to the identity of

### Outcome/Findings

The test re-test reliability fluctuated greatly between tasks ranging from .20 to .76. Individual task scores are not recommended for use in determining eligibility for services or in documenting change over time.
|---|---|---|---|
| 1) **ETCH-M**  
2) Bruinin ks-Oseretsky Test of Motor Proficiency (BOTMP)  
3) In-Hand Manipulation Test (IHM)  
4) Developmental Test of Visual-Motor Integration, Revised (VMI)  
5) Test of Visual-Perceptual Skills-Non-Motor (TVPS)  
6) Finger Identification (FI)  
7) Motor Accuracy (MAC) Steadiness Test | Validity Research Article Available at Region X | To determine the relationship between ETCH-M scores and teacher ratings of handwriting performance. | 3 |
| 69 first graders, 32 boys and 37 girls from 40 schools. Mean age was 82 ± 3.6 months. All were in regular education programs. Eight students were receiving resource help and four were receiving OT. | Each student was administered eight tests individually in an enclosed room at a local hospital by an experienced pediatric OT for 2 hours with a rest break. The teacher completed a handwriting scale for each student. | The correlation between the teachers’ ratings and the ETCH-M scores was (r=0.40-0.45; p<0.05). Which means the teacher’s ratings and the ETCH scores were similar, supporting the validity of the ETCH-M. |

|---|---|---|---|
| ETC – C  
Research Article Available at Region X | Concurrent validity is the ability to distinguish between groups taking the test. | 3 |
| 101 fourth graders from 4 elementary schools in a large suburban school district. Consent was obtained. Age of participants 9 yrs 0 mths to 10yrs 7mths (M=9 yrs 2.3 mths). 94% regular educational, 5% certified speech & language impaired, 1% hearing | Used the ETCH-C and the Cursive Practice and Review Worksheet. Two samples of handwriting were collected from each student during a 4-week period in fall 1999. Each student was administered the ETCH-C and later the Cursive Practice and Review worksheet. The ETCH | Concurrent validity is weakly supported with coefficients .61 to .65. |
impaired & learning disabled. No students were receiving OT at time of study.

was administered by the researcher while the Cursive Practice was administered in the classroom by the teacher. The worksheets were graded A, B, C, or unsatisfactory by the teacher to then be compared to the ETCH-C for validity.

| ETCH-M | Ecological Validity Research Article Available at Region X | Does the ETCH measurements of the handwriting performance of a student match/correlate to the teacher’s perceived performance? | 3 | 45 students who were identified by their teachers as having handwriting difficulties. There were 30 boys and 15 girls, ages 6 years 2 months to 7 years 11 months. The student was tested individually in a quiet location. The teacher was given a questionnaire the same day or day after the student was tested. The questionnaire contained 6 questions on handwriting tasks to determine the child’s overall and tasks specific performance compared to his or her peers. The tasks were comparable to the ETCH tasks. | The correlation coefficients between the teacher’s overall perception of the student’s performance and the ETCH total word, letter, and number legibility were low and not significant. There was no significant relationship between each ETCH subtask and the teacher’s perception of the task. NOTE: The ETCH may not be able to distinguish between poor and severely poor handwriting performance thus leading to the results. |

| ETCH - M | Research Article Available at Region X | Explains Diekema, Deitz, & Amundson study, its relevance to clinical practice | 5 | NA | NA | Explains Diekema, Deitz, & Amundson study and its relevance to clinical practice |
**ETCH**

**EBPX Summary**

*Summary of the EBPX team on the collective evidence reviewed.*

The interrater reliability of the ETCH has a wide range of 0.63 to 0.91 for experienced administrators (Tomchek & Schneck, 2007).

The test re-test reliability had a range of .63 to .71.

Both the interrater and test re-test reliability coefficients are not strong. The fact that the ETCH is a holistic test (based on perceptual judgment of the scorer) rather than error referenced (objective recording of deviations from a provided sample) has an impact on the reliability of the assessment. These figures suggest the reliability of the scores may be questionable from time to time and from administrator to administrator.

The ETCH has support for face and content validity. The results of Koziatek and Powell’s study (2002) showed that the ETCH-C does have a degree of concurrent validity with how teachers assign handwriting grades. The assessment was able to distinguish between teacher grades of satisfactory vs. unsatisfactory for letter legibility successfully 73% to 82%.

**EBPX Strength and Impact Summary**

*Interpretation of the collective evidence reviewed by the EBPX team.*

The ETCH has good face and content validity. This means the evidence suggests the ETCH does test handwriting legibility. It is also a good tool in that it considers a wide range of tasks and testing domains.

The ETCH lacks good reliability. This means the scores between different testers and different times can be very different. When the ETCH is used, it needs to be...
scored by the same individual each time the test is given to the child. Also therapists must be cautious about using the ETCH to measure change in handwriting performance. A change in ETCH score from pre to post treatment, particularly a relatively small change, may be due to the poor reliability of the test and not due to a change in the child’s actual performance.

There is SOME EVIDENCE THAT THIS ASSESSMENT IS EFFECTIVE. Therapists are justified in giving some consideration to the use of this assessment.

**Children’s Handwriting Evaluation Scale (CHES)**

To purchase this evaluation you can call the Author, Jody Phelps, at (214)-366-3667

**CHES**

**Assessment Description:**

Phelps and Stempel (1987, 1988) in the Child Development Division at Texas Scottish Rite Hospital in Dallas, Texas developed the Children’s Handwriting Evaluation Scale (CHES). There are two versions, manuscript and cursive. The manuscript version is for grades 1 and 2. The cursive version is for grades 3 through 8. The CHES is designed to remediate handwriting problems and to prevent further difficulties (Feder, 2003).

This is a norm referenced for the handwriting speed assessment. The CHES-M asks students to copy 2 sentences, consisting of 57 letters including all letters of the alphabet except i, q, v, x, and z in a 2 minute period (Feder, 2003). The results are scored for speed using a rate table and quality. The quality scoring looks at letter form, spacing, rhythm, and general appearance which have loosely define objective criteria.

**Evidence Table**

Contains appraisals of evidence reviewed.

**Key to Level of Evidence**

*(Level of evidence may be adjusted downward if study has poor rigor.)*

<table>
<thead>
<tr>
<th>Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Evidence</strong></td>
<td>Systematic Reviews and meta-analyses</td>
<td>Randomized Control Trials (RCT)</td>
<td>Quasi-experimental and Comparative studies</td>
<td>Correlation and Non-experimental studies</td>
<td>Descriptive studies &amp; Expert Opinion articles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citation</th>
<th>Handwriting Assessments</th>
<th>Type of Evidence &amp; Access</th>
<th>Description of Evidence/ Type of study</th>
<th>Level of Evidence</th>
<th>Description of Population</th>
<th>Description of Intervention</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phelps, J. &amp; Stempel, L. (1987). Handwriting: Evolution and CHES-C</td>
<td>Reliability: Available at Region X Intrarater</td>
<td>Interrater</td>
<td>3</td>
<td>1,365 children ranging from grades three to eight in Random groups of 25 papers were drawn</td>
<td>Reliability of single rater r=0.64–0.82. Reliability mean of four</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dallas county provided handwriting samples from each grade level and then were scored by teachers and therapists in the hospital’s dyslexia lab. Each paper was evaluated by 3 graders and both authors r=0.88-0.94. Standard error of measurement based on a single rater r=0.52-0.68.

| Phelps, J. & Stempel, L. (1988). The children’s handwriting evaluation scale for manuscript writing. Reading Improvement, 25(4), 247-254. | CHES-M | Reliability: Research Article Available at Region X http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_tbs=ecs_retrieve&_&ERICExtsSearchSearchValue_0=EJ386986&ERICExtSearchSearchType_0=no&accno=EJ386986 | Interrater Intrarater | 3 | 643 first and second grade students in Dallas county. | 2 professional examiners assisted by experienced volunteers administering the tests. Two professional members of the Child Development staff were asked to independently score 25 papers randomly selected from each grade. The authors’ scores were compared to the Child Development staff scores. | Grade 1 scores: Single rater r=0.81 Mean of 3 raters r=0.93 Grade 2 scores: Single rater r=0.65 Mean of 3 raters r=0.85. | Judgments of handwriting coincided 76% of the time for Grade 1 and 72% for Grade 2. |

EBPX Summary
*Summary of the EBPX team on the collective evidence reviewed.*

The CHES has low to moderate evidence for interrater ($r=0.85-0.95$) and intrarater ($r=0.64-0.82$) reliability.

There is no evidence on re-test reliability or validity.

CHES is a quick and easy way to assess a child’s handwriting performance. It only takes 2 minutes to administer and 3 to 7 minutes to score. This would be a good screening tool when evaluating a child’s handwriting performance.

EBPX Strength and Impact Summary
*Interpretation of the collective evidence reviewed by the EBPX team.*

There is suggestive evidence that the Children’s Handwriting Evaluation Scale (CHES) may be effective in assessing a child’s handwriting performance.

There would be difficulty to measure the change and progress of a child’s handwriting performance with the CHES due to the lack of test-retest reliability.

There is no evidence at this time that the CHES is able to discriminate between poor and good writers, only the extremes.

CHES is a quick and easy way to assess a child’s handwriting performance. It only takes 2 minutes to administer and 3 to 7 minutes to score. This would be a good screening tool when evaluating a child’s handwriting performance.

**Minnesota Handwriting Test (MHT) aka Minnesota Handwriting Assessment (MHA)**

**MHT/MHA**

**Assessment Description:**

Judith E. Reisman developed the Minnesota Handwriting Test (MHT). This tool is norm referenced and measures changes in handwriting performance of first and second graders. MHT tests near point copying of manuscript and the speed of the handwriting performance. Students are requested to copy the scrambled words of the sentence “the quick brown fox jumped over lazy dogs” for 2.5 minutes.

Scores are obtained for quality and rate using five categories. The five categories are legibility, form, alignment, size and spacing. If the sample is not legible, the scoring is stopped and does not continue to the other four categories. If the sample is legible, scoring continues. Form measures the quality of the letters. Alignment is the placement of the letters in relation to the baseline. Size is the how the letters relate to the midline. Spacing encompasses both letter and word spacing. The four categories of form, alignment, size and spacing have specific measurements and criteria for scoring.

**Evidence Table**

*Contains appraisals of evidence reviewed.*

**Key to Level of Evidence**

*(Level of evidence may be adjusted downward if study has poor rigor.)*

<table>
<thead>
<tr>
<th>Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Evidence</td>
<td>Systematic Reviews and meta-analyses</td>
<td>Randomized Control Trials (RCT)</td>
<td>Quasi-experimental and Comparative studies</td>
<td>Correlation and Non-experimental studies</td>
<td>Descriptive studies &amp; Expert Opinion articles</td>
</tr>
<tr>
<td>Citation</td>
<td>Handwriting Assessments</td>
<td>Type of Evidence &amp; Access</td>
<td>Description of Evidence/Type of study</td>
<td>Level of Evidence</td>
<td>Description of Population</td>
</tr>
</tbody>
</table>

Page 10 of 13
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minnesota Handwriting Test (MHT)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Remediation of Handwriting Problems</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The differences between the groups was significant (F=218.7, p<.0001). The MHT scores correlated to the teachers perceived need for intervention.
<table>
<thead>
<tr>
<th>MHT</th>
<th>Reliability</th>
<th>Interrater</th>
<th>6 raters were trained, then scored 10 of the students samples independently. The scores were then compared to the author’s scoring, more training of raters was conducted. The process was repeated a second and third time.</th>
<th>The reliability between the scorer and the 6 raters increased as the raters became more experienced. Experienced raters r=0.99 for total, while categories ranged from r=0.90 to 0.99. Inexperienced raters r=0.975 for total, while categories ranged from r=0.87 to .98. Strong interrater reliability.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrarater</td>
<td>3</td>
<td>Author, an experienced rater, and an inexperienced rater</td>
<td>Scored 20 samples and rescored them 5 to 7 days later.</td>
<td>Author achieved 98.5%, experienced rater scored 98.8%, and the inexperienced rater scored 96.4%. All correlations for category scoring was r=.93 to .99 and total scoring r=.99. Strong intrarater reliability</td>
</tr>
<tr>
<td>Test retest</td>
<td>3</td>
<td>Convenience sample of 99 second graders in three schools.</td>
<td>Each student completed the MHT twice with a time interval of 5 to 7 days between tests. Both tests were administered by teachers or therapists to each class. The tests were scored by the author and her research assistances.</td>
<td>Correlation coefficients for total accuracy r=0.72 with a between-school range of r=0.58-0.94. Correlation coefficients for rate scores r=0.50 with a between-school range of r=0.47-0.67. Low to moderate test retest reliability.</td>
</tr>
</tbody>
</table>
EBPX Summary
Summary of the EBPX team on the collective evidence reviewed.

The interrater reliability of the Minnesota Handwriting Test has a strong range of 0.87-0.98.

The intrarater reliability is even stronger with a range of 0.93 – 0.99.

The test retest reliability has a wide range from poor to moderate of 0.58 – 0.94.

The MHT has been shown to have good intra-rater and inter-rater reliability with content validity. The test re-test reliability was low to moderate questioning the MHT as a reliable tool to measure change over time.

EBPX Strength and Impact Summary
Interpretation of the collective evidence reviewed by the EBPX team.

There is SUGGESTIVE EVIDENCE that the Minnesota Handwriting Assessment (MHA) also known as Minnesota Handwriting Test (MHT) may be effective in assessing a child’s handwriting performance.

There is little evidence that the MHA may be able to measure change and progress in a child’s handwriting performance, due to the wide range of test retest scores. There is some discrimination but the accuracy of the change may be questionable.