Assessing Writing Skills Using Correct–Incorrect Word Sequences: A National Study

Kristina C. Breaux, PhD & Frances E. Frey, MA
WIAT-III® Development Team
Clinical Test Development, Pearson

Abstract
A study of Correct minus Incorrect Word Sequences (CIWS) as a measure of writing skills was conducted with over 1,000 students, including groups of exceptional learners. The study investigated reliability, developmental trends, demographic bias, and how well CIWS discriminates between students with and without learning disabilities and other exceptionalities. Results indicate that CIWS is a relatively unbiased measure with strong inter-scorer reliability. An expected developmental progression was revealed, and the clinical studies indicated that CIWS is an effective measure for differentiating between groups. Participants will learn how to use CIWS to score writing samples, what CIWS measures, how well CIWS discriminates between clinical groups, the strength of its technical properties, and its application to progress monitoring and determining eligibility.

Introduction

• Correct minus Incorrect Word Sequences (CIWS) has traditionally been used as a curriculum–based measure of written grammar and mechanics.
• Previous Research (Espin et al., 2000; Espin, De La Paz, et al., 2008; Espin, Wallace, et al., 2008)
  o CIWS moderately correlated with teacher ratings of student writing samples and performance on state/district writing tests.
  o CIWS is a valid measure of narrative and expository essay writing in the lower and upper grades.
  o CIWS validity coefficients typically range from .6 – .8.
  o CIWS has relatively weaker validity for high performing students (.56 – .6).
  o CIWS inter-scorer agreement is consistently reported between 88% and 92%.
  o CIWS is sensitive to change over time, and an effective measure for progress monitoring.
• Large-scale studies of CIWS have not been conducted, and many studies have not adequately described and differentiated their clinical samples.

CIWS Scoring Rules

• CWS: A Correct Word Sequence is two adjacent words that are correctly spelled, capitalized, and punctuated, and grammatically and semantically acceptable within the context of the sentence.
  A caret (^) is used to mark each CWS.
  ▪ If the first word of a sentence is capitalized, always mark a CWS in front of the word.
  ▪ If correct ending punctuation follows the last word of a sentence, always mark a CWS after the word.
• IWS: An Incorrect Word Sequence is two adjacent words that do not qualify as a CWS.
  A dot (.) is used to mark each IWS.
• CIWS = CWS – IWS: To calculate the CIWS score, sum CWS and IWS separately, and subtract IWS from CWS.

Scoring Example

<table>
<thead>
<tr>
<th>CWS</th>
<th>IWS</th>
<th>Scoring Example: Item 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>Write about your favorite game. Include at least 3 reasons why you like it.</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>&quot;Hide and Seek&quot; is a good game. I like playing it.</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>That game at home because there is lots of places to hide. You can play it anywhere.</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Reason I like &quot;Hide and Seek&quot; is that you can pretend to play that game.</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>You can pretend to play that game.</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Hide away from your parents.</td>
</tr>
</tbody>
</table>

Essay Scoring Example: Item 1

Write about your favorite game. Include at least 3 reasons why you like it.

"Hide and Seek" is a good game. I like playing it. That game at home because there is lots of places to hide. You can play it anywhere. Reason I like "Hide and Seek" is that you can pretend to play that game. Hide away from your parents.
**Research Design**

**Sample**
- Nationally stratified and representative sample.
- 1,000 students in grades 3 – 12 (100 per grade) from the standardization sample of the *Wechsler Individual Achievement Test – Third Edition* (WIAT–III®).
- 5 special studies:
  - Academically Gifted and Talented (GT; n = 61)
  - Mild Intellectual Disability (Mild Mental Retardation) (MID; n = 35)
  - Learning Disorder in Writing (LD-W; n = 44); comorbid LD–R acceptable
  - Learning Disorder in Reading (LD-R; n = 78); comorbid LD–W acceptable
  - Learning Disorder in Math (LD-M; n = 51); comorbid LD–R acceptable

**Analyses**
- Inter–Scorer Reliability
  - Correlation: A Pearson correlation between pairs of CIWS raw scores by independent scorers.
  - Percent Agreement: For each pair of raw scores, divided the lower score by the higher score and multiplied by 100.
  - Scale Gradient: Examined the degree of change in standard scores as a result of change in CIWS raw scores.
- Development Trends
- Demographic Differences
- Clinical Utility

**Results**

**Inter-scorer Reliability**
- Correlation: 99%
- Percent Agreement: 90%
- Scale Gradient: On average, a change of 3 raw score points resulted in a change of 1 std. score point.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Raw score change for 1 standard score point change</th>
<th>Raw score change for 1 standard deviation change</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 – 4</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>5 – 9</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>10 – 12</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>Avg</td>
<td>3</td>
<td>45</td>
</tr>
</tbody>
</table>

**Development Trends**
Mean CIWS raw scores generally increase as grade level increases.
Demographic Differences

The Nation’s Report Card (NCES, 2003) reported the following statistically significant results from the NAEP writing assessment:

- Females scored higher than males (grades 4, 8, 12).
- White & Asian/Pacific Islander students scored higher than Black and Hispanic students (grades 4, 8, 12).
- Hispanic students scored higher than Black students in grade 12.

This study found similar statistically significant results for CIWS performance:

- Females scored higher than males overall (and in each grade except 9 and 11).
- White and Other students scored higher than Black and Hispanic students overall, but almost never by grade (White students outperformed Black students in grades 5 and 7).
- Hispanic students scored higher than Black students overall, but not by grade.

<table>
<thead>
<tr>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Mean</td>
</tr>
<tr>
<td>CIWS Std. Score</td>
<td>103.51</td>
</tr>
</tbody>
</table>

**p < .01.

Comparison | Mean Difference | Std. Error | Sig. | d
(x) | (y) | (x-y) |
White | Black | 8.14 | 1.39 | <.001** | .47 |
Hispanic | 3.76 | 1.23 | .013* | .20 |
Other | Black | 10.50 | 2.23 | <.001** | .67 |
Hispanic | 6.12 | 2.14 | .023* | .36 |
White | 2.36 | 1.94 | .615 | .14 |
Black | Hispanic | -4.38 | 1.66 | .042* | -2.7 |

*p < .05. **p < .01.

Clinical Utility

Matched t-tests revealed significant differences between the NA group and each of the special groups; this finding supports previous research (Espin, Wallace, et al., 2008).

<table>
<thead>
<tr>
<th>Comparison</th>
<th>n</th>
<th>Clinical</th>
<th>Mean</th>
<th>SD</th>
<th>Normal</th>
<th>Mean</th>
<th>SD</th>
<th>diff</th>
<th>t</th>
<th>Sig.</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD-W / NA</td>
<td>44</td>
<td>85.3</td>
<td>9.7</td>
<td>102.9</td>
<td>14.6</td>
<td>17.66</td>
<td>6.40</td>
<td>&lt;.001**</td>
<td>1.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD-R / NA</td>
<td>78</td>
<td>87.6</td>
<td>8.9</td>
<td>99.5</td>
<td>11.7</td>
<td>11.83</td>
<td>7.04</td>
<td>&lt;.001**</td>
<td>1.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD-M / NA</td>
<td>51</td>
<td>88.0</td>
<td>12.3</td>
<td>102.7</td>
<td>15.3</td>
<td>14.67</td>
<td>5.46</td>
<td>&lt;.001**</td>
<td>1.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID / NA</td>
<td>35</td>
<td>80.3</td>
<td>7.2</td>
<td>99.5</td>
<td>13.7</td>
<td>19.17</td>
<td>6.50</td>
<td>&lt;.001**</td>
<td>1.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT / NA</td>
<td>61</td>
<td>112.6</td>
<td>15.6</td>
<td>105.5</td>
<td>14.1</td>
<td>-7.16</td>
<td>-2.81</td>
<td>&lt;.01**</td>
<td>-4.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p < .01
Clinical Studies

Group Comparisons

Mean CIWS Standard Scores

Clinical Group

Normally Achieving Group

Conclusions

This research supports CIWS as a valid and reliable measure of written expression (grammar and mechanics) for use with a diverse student population.

- **Inter-scorer Reliability**: CIWS is a robust and highly reliable measure that is not sensitive to small amounts of variance in scoring.

- **Development Trends**: CIWS means and standard deviations increase expectedly with grade level.

- **Demographic Bias**: CIWS is an unbiased measure that is sensitive to validated gender and ethnic differences in writing ability. Such differences may be explained by cultural, environmental, and developmental factors.

- **Clinical Utility**: CIWS differentiates between NA students and LD–W, LD–R, LD–M, MID, and GT groups.

Applications

- **Comprehensive Evaluation**: CIWS can be converted to a norm–referenced score.
  - The WIAT–III® includes CIWS as a measure of written expression, yielding a supplemental standard score to assist practitioners in determining eligibility, diagnosing writing disorders, and/or planning intervention.

- **Progress Monitoring**: The WIAT–III® yields percentile ranks for CIWS, which can be used in conjunction with CIWS raw scores to track progress over time.

References


