



# How Valid Is Curriculum-Based Measurement in Written Expression?

Tanya L. Eckert, Adrea J. Truckenmiller, Jennifer L. Rymanowski,  
Jennifer L. Koehler, Elizabeth A. Koenig, and Bridget O. Hier

Syracuse University, Department of Psychology



## Introduction

Currently, there is a great need for appropriate assessment and effective intervention in written expression. The National Center for Education Statistics (2003) reported that only 28% of fourth-grade students were writing at the level expected of them. In order to effectively target elementary-aged students' writing competence and formatively evaluate their progress, assessment measures must be well-documented as valid for their intended use (AERA, APA, & NCME, 1999). Curriculum-based measurement in written expression (CBM-WE) is one formative evaluation tool that has been recommended (Deno, Marston, & Mirkin, 1981). Recent reviews of the technical adequacy of CBM-WE have revealed limited criterion validity evidence (McMaster & Espin, 2007), with most studies focusing on the association between CBM-WE and prior editions of the Test of Written Language. No studies have used other standardized measures of writing.

## Purpose

The purpose of the present study was to examine the criterion validity of CBM-WE in relation to commonly used standardized measures of written expression including: (a) the spontaneous writing subtest of the Test of Written Language – Third Edition (TOWL-3; Hammill & Larsen, 1996); (b) the paragraph writing portion of the Wechsler Individual Achievement Test - Second Edition (WIAT-II; Psychological Corporation, 2002); and (c) the Broad Written Language Cluster of the Woodcock Johnson Test of Achievement-Third Edition (Woodcock, McGrew, & Mather, 2001). The relationship between these measures and commonly-used scoring metrics obtained from CBM-WE probes, including total words written (TWW), (b) number of correct writing sequences (CWS), and number of correctly spelled words (CSW), was examined.

## Participants and Setting

- A total of 425 third-grade students in schools that served students in Kindergarten through grade six participated in the study.
- More participants were female (53.2%) than male (46.8%).
- The majority of participants self-identified their race as either Black/African American (50.1%) or White (38.1%).
- The average age was 8.02 years (SD = .46).
- Most of the participants were eligible for free or reduced-price lunch (84%)

## Procedures

### Materials

- CBM-WE probes (McMaster & Campbell, 2006; Shinn, 1998)
- Spontaneous portion of the TOWL-3 (Hammill & Larsen, 1996)
- Paragraph writing portion of the WIAT-II (Psychological Corporation, 2002)
- Broad Written Language Cluster of the WJ-III (Woodcock et al., 2001)

Week 1	Week 2	Week 3
TOWL-3 (15 min.)	WIAT-II (10 min.)	WJ-III (10 min.)
CBM-WE (3 min.)	CBM-WE (3 min.)	CBM-WE (3 min.)

- Missing data were obtained for 2.3% of TOWL-3, 3.5% of the CBM-WE probes, 55.9% of the WIAT-II, and 89.2% of the WJ-III.
- Responses were score able on 69.7% of TOWL-3, 100% of the WIAT-II, 100% of the WJ-III, and 100% of the CBM-WE probes.

## Results

Each CBM-WE metric was correlated with the age-based standardized scores of the TOWL-3, WIAT-II, and WJ-III. The descriptive results (see Table 1) and the correlational results (see Table 2) are presented below.

Table 1. Descriptive findings

Metric	Mean	(SD)
CBM-WE (TWW)	26.37	(12.58)
CBM-WE (CWS)	24.58	(12.43)
CBM-WE (CSW)	23.58	(12.04)
TOWL-3	69.60	(48.23)
WIAT-II	102.23	(11.72)
WJ-III	75.83	(29.94)

Table 2. Association between CBM-WE and standardized measures of written expression

Metric	TOWL-3	WIAT-II	WJ-III
CBM-WE (TWW)	.28**	.66**	.47**
CBM-WE (CWS)	.30**	-----	.58**
CBM-WE (CSW)	.35**	.64**	.66**

\*\* $p \leq .01$

## Discussion

- For elementary-aged students, this study reveals that correct writing sequences (CWS) and correctly spelled words (CSW) are more valid metrics than total words written (TWW).
- The results also suggest that CBM-WE is more strongly associated with the WIAT-II and WJ-III standardized measures than the TOWL-3. This may be related to the moderate percentage of TOWL-3 responses (30.3%) that could not be included in the analysis due to the scoring criterion established by this measure.