



Predictive validity of curriculum-based measurement and teacher ratings of academic achievement[☆]



Ryan J. Kettler^{a,*}, Craig A. Albers^b

^a Rutgers, The State University of New Jersey, USA

^b University of Wisconsin – Madison, USA

ARTICLE INFO

Article history:

Received 23 May 2011

Received in revised form 25 February 2013

Accepted 26 February 2013

Keywords:

Curriculum-based measurement

Performance screening guides

Predictive validity

Sensitivity

Teacher ratings

Universal screening

ABSTRACT

Two alternative universal screening approaches to identify students with early learning difficulties were examined, along with a combination of these approaches. These approaches, consisting of (a) curriculum-based measurement (CBM) and (b) teacher ratings using Performance Screening Guides (PSGs), served as predictors of achievement tests in reading and mathematics. Participants included 413 students in grades 1, 2, and 3 in Tennessee ($n = 118$) and Wisconsin ($n = 295$) who were divided into six subsamples defined by grade and state. Reading and mathematics achievement tests with established psychometric properties were used as criteria within a concurrent and predictive validity framework. Across both achievement areas, CBM probes shared more variance with criterion measures than did teacher ratings, although teacher ratings added incremental validity among most subsamples. PSGs tended to be more accurate for identifying students in need of assistance at a 1-month interval, whereas CBM probes were more accurate at a 6-month interval. Teachers indicated that (a) false negatives are more problematic than are false positives, (b) both screening methods are useful for identifying early learning difficulties, and (c) both screening methods are useful for identifying students in need of interventions. Collectively, these findings suggest that the two types of measures, when used together, yield valuable information about students who need assistance in reading and mathematics.

© 2013 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Educators and parents alike often look to school psychologists as interpreters of assessment results, as well as experts who will submit an informed opinion about whether a student is in need of academic assistance, whether it be proactive early intervention services or special educational services. Although assessment procedures must occur as part of determining a specific learning disability classification, assessment of risk status for learning difficulties is designed to occur at an earlier stage of skill development so that proactive and early intervention approaches can be implemented. Interventions also should be attempted before a student fails academically to the degree that a special education evaluation is warranted. Federal legislation (i.e., [No Child Left Behind Act \[NCLB\], 2001](#)) emphasized this need for an increased use of universal screening instruments to identify students at risk for

[☆] The current study was implemented as part of Early Identification of Students with Learning Difficulties: A Comparison of Methods, a project funded by the Early Career Awards Program of the Society for the Study of School Psychology. The authors of the current manuscript were the co-principal investigators for that project. They wish to thank Thomas R. Kratochwill for his service as mentor on the project. The positions and opinions expressed in this article are those solely of the authors.

* Corresponding author at: Graduate School of Applied and Professional Psychology, Rutgers, The State University of New Jersey, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA. Tel.: +1 848 445 3960.

E-mail address: r.j.kettler@rutgers.edu (R.J. Kettler).

ACTION EDITOR: Robin Coddling.

academic failure and in need of proactive early interventions. Thus, universal screening instruments need to be efficient assessment tools that can be used to determine the likelihood that a student will experience learning difficulties. Consequently, the development of universal screening instruments and the validation of their scores are extremely important within education, because early detection of children at risk for learning difficulties will facilitate the replacement of reactionary responses with preventative and early intervention approaches (e.g., Albers, Glover, & Kratochwill, 2007a, 2007b; Glover & Albers, 2007).

Curriculum-based measurement (CBM) and teacher ratings of students' performance are two methods of screening that hold promise for identifying students at risk for academic difficulties. CBM probes are brief tests of foundational academic skills, such as oral reading fluency and mathematics calculation, which are designed to represent the types of learning activities that students experience in the classroom (e.g., Fuchs & Deno, 1991; Fuchs & Fuchs, 2000; Reschly, Busch, Betts, Deno, & Long, 2009). Teacher rating scales consist of teacher judgments of students based on multiple weeks and months of interaction, which can incorporate characteristics related to academic achievement (e.g., academic enablers and opportunity to learn) that are not directly testable (Kettler, Elliott, Davies, & Griffin, 2012; for a more thorough comparison between CBM and teacher ratings for screening, the reader is directed to Elliott, Huai, & Roach, 2007). The use of CBMs and teacher ratings of students' performance have been the focus of a great deal of research in the area of universal screening, but the results have been difficult to compare because analyses typically have been completed using different criterion measures on different participant samples. The current study was designed to compare the individual and collective prediction of CBM probes and teacher ratings for use as universal screening measures to identify students who are in need of academic assistance.

1.1. CBM probes as predictors of achievement

CBM is based on procedures used to characterize student performance in basic content areas. CBM focuses on broad objectives, allows for retention and generalization, and specifies standard measurement and evaluation procedures. Three core design features of CBM include (a) constructing the materials or probes, (b) administering and scoring, and (c) organizing data and decision making (Hintz, Christ, & Methe, 2006). Much evidence has been collected on the validity of inferences from CBM probes based on relations with other measures of achievement. Correlation coefficients range from large to nearly perfect¹ between CBM reading fluency measures and criterion-referenced basal reading mastery tests, and from medium to nearly perfect between CBM probes and norm-referenced tests (e.g., Marston, 1989; Reschly et al., 2009; Shinn, Good, Knutson, Tilly, & Collins, 1992). With regard to measurement of growth, Fuchs, Deno, and Mirkin (1984) demonstrated a statistically significant relation between oral reading improvements over time and improvements on standardized measures of reading. McGlinchey and Hixson (2004) investigated the predictive validity of CBM probes on fourth grade students' performance on the Michigan Educational Assessment Program's (MEAP; Michigan State Board of Education, 1999) reading assessment, yielding a correlation of .67. More recently, Reschly et al. (2009) conducted a meta-analysis of the correlational evidence between reading CBM and other standardized tests of reading achievement; this review of 289 studies indicated a mean correlation between reading CBM probes and reading achievement of .68.

CBM probes in mathematics have also been shown to be effective for predicting achievement. Clarke and Shinn (2004) found large to very large correlations (i.e., $r = .60$ to $.79$) between mathematics CBM probes and the Woodcock–Johnson Tests of Achievement, Revised Applied Problems test (Woodcock & Mather, 1989), and Chard et al. (2005) found a large correlation ($r = .53$) between mathematics CBM probes and the Number Knowledge Test (Okamoto & Case, 1996). Criterion validity coefficients between the Monitoring Basic Skills Progress (MBSP) Basic Math Computation (Fuchs, Hamlett, & Fuchs, 1999) CBM probes and the Math Computation Test (Fuchs, Fuchs, Hamlett, & Stecker, 1991) ranged from .82 to .83 across all grade levels, whereas correlations between the MBSP and the Stanford Achievement Test (Gardner, Rudman, Karlsen, & Merwin, 1982) ranged from .66 to .67.

1.2. Teacher ratings as predictors of achievement

Research indicates that teacher judgments can be accurate predictors of students' needs for academic assistance. Gerber and Semmel (1984) suggested valuing teacher judgments because teachers have daily contacts with students in a meaningful context in which to evaluate academic performance. Gresham, Reschly, and Carey (1987) indicated that teachers correctly identified 96% of students diagnosed with a learning disability, and Gresham, MacMillan, and Bocian (1997) reported that teachers correctly identified 91% of students with learning disabilities, 95% of students exhibiting low achievement, and 100% of students with low IQs.

Several studies have been conducted using the Brief Academic Competence Evaluation Screening System (BACESS; Elliott, DiPerna, & Huai, 2003), which includes a set of nomination rubrics designed for universal screening (Kettler, 2007; Kettler & Elliott, 2010; Kettler, Elliott, & Albers, 2008) of academic problems. These studies have incorporated state achievement proficiency tests as criterion measures. On average, BACESS scores have had high specificity² (.82), moderate positive predictive value (PPV; .76) and negative predictive value (NPV; .66), but low sensitivity (.57). Kettler et al. (2012) conducted similar research using nomination rubrics (i.e., Performance Screening Guides; PSGs) from the Social Skills Improvement System (SSiS;

¹ We recognize that there is no standard heuristic for providing nominal labels for r values. We used the following system introduced by Cohen (1992), extended by Hopkins (2001), and applied in educational assessment by Kettler et al. (2010): $r = .00 \leq$ nonexistent $< .10 \leq$ small $< .30 \leq$ medium $< .50 \leq$ large $< .70 \leq$ very large $< .90 \leq$ nearly perfect.

² We used the following system from Kettler and Feeney-Kettler (2011) to provide nominal labels for conditional probability indices: $.00 \leq$ very low $< .40 \leq$ low $< .60 \leq$ moderate $< .80 \leq$ high < 1.00 .

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات