Using Curriculum-Based Measurement to Prevent Failure and Assess Learning in the Content Areas

Todd W. Busch
University of South Carolina

Christine A. Espin
University of Minnesota

Many students continue to experience reading difficulties into their secondary-school years. Difficulties include poor word decoding, limited vocabulary knowledge, and deficits in working memory capacity. Such difficulties limit the ability of students to acquire new knowledge and be successful in content-area classes. In this article, we describe a curriculum-based measure, vocabulary matching, that teachers can use to assess and monitor the performance of students with reading difficulties in the content areas. Vocabulary matching, has been found to be a valid and reliable measure of performance and progress in content areas. A case study is provided to illustrate the use of vocabulary matching in content-area instruction.

At the secondary-school level, students are expected to read to learn rather than to learn to read (Alley & Deshler, 1979). Unfortunately, many students reach secondary school not yet proficient in reading. These students experience difficulties in many areas of reading, but three are especially relevant in terms of reading and understanding content-area text: fluent word decoding (Young & Bowers, 1995), vocabulary knowledge (Baumann & Kameenui, 1991), and working memory capacity (Swanson, Cochran, & Ewers, 1990).

Perhaps the most well-recognized area of difficulty for students with reading problems is poor word decoding fluency. Students with reading disabilities (RD) tend to be slower and less accurate word decoders than average readers (Young & Bowers, 1995). LaBerge and Samuels (1974) hypothesized that this slow and inaccurate word decoding forces students to allocate limited cognitive resources to decoding text rather than to constructing meaning from text. The comprehension difficulties caused by poor word decoding fluency are compounded when students are required to read content-specific texts in social studies and science, which are often less well structured than narrative texts, making comprehension particularly difficult (Linderholm et al., 2000).

A second area of reading difficulty for students with RD is limited vocabulary knowledge (i.e., understanding the meaning of words once they have been decoded). The relation between vocabulary knowledge and reading comprehension is viewed as reciprocal (Stanovich, 1986): Strong beginning reading skills facilitate vocabulary growth — vocabulary growth, in turn, contributes to improved reading comprehension. It is not surprising, then, that students with RD have less extensive vocabularies than students without RD, or that the interdependence between vocabulary knowledge and reading comprehension increases as students get older (Simmons & Kameenui, 1990). Limited vocabulary knowledge combined with poor word decoding fluency puts students with RD at risk for failure in content-area classes.
where students are to read and understand terminology-laden text.

A third area of difficulty for secondary-school students with RD is poor working memory capacity (i.e., the capacity to store and process information). In a study of 41 undergraduate students, Daneman and Carpenter (1980) found that individuals with a smaller reading span (used as an indicator of working memory) scored lower on comprehension measures than individuals with a larger reading span. The authors hypothesized that less skilled readers needed to devote more working memory capacity to the process of reading, and thus were limited in their ability to store information from text for later recall. Other support for working memory capacity deficits in students with RD has come from Torgesen (1988), Swanson et al. (1990), and Gettinger (1991), all of whom found working memory capacity deficits for students with RD.

In sum, areas of reading difficulties for secondary students with RD include poor word decoding fluency, limited vocabulary knowledge, and insufficient working memory capacity. These difficulties limit students’ capacity to comprehend text, especially content-specific text. Limitations in the ability to read and understand content-specific text put students with RD at risk for failure in their content area classes. High failure rates, in turn, are associated with high dropout rates (Donahoe & Zigmond, 1990; Wagner, 1990). Secondary-school teachers must act to prevent student failure in their content-area classes, but what can they do?

**Preventing Failure and Assessing Learning in the Content Areas**

One approach to failure prevention is early identification and monitoring of at-risk students in content-area classes. To identify and monitor at-risk students, teachers must have at their disposal a tool that is valid with respect to student performance and learning in the content areas. Given the reading skills required of students in the content areas, a valid tool would require word decoding fluency, vocabulary knowledge, and working memory capacity. To assess learning over time, the tool also must be appropriate for use on a repeated basis: Thus, it must be reliable, of short duration, and easy to administer and score.

Curriculum-based measurement (CBM) has been used to assess the performance and progress of students in basic skill areas. The measures used as a part of CBM are valid indicators of general student performance in reading, written expression, spelling, and mathematics (see Deno, 1985). In addition, the measures are reliable, of short duration, and easy to administer and score. CBM is designed to encourage implementation of interventions before a student fails. When using a CBM approach, teachers administer probes of short duration to students on a weekly basis and graph the students’ performance each week. Teachers subsequently use the graphed data to evaluate student performance and determine the success of their instructional interventions. When students are not progressing, teachers change instruction, and then examine subsequent data to evaluate the effects of that change.

The majority of CBM research has been conducted at the elementary-school level in basic skills areas. Recently, however, CBM research has been extended to the secondary-school level, where it has focused on basic skills areas such as reading (Espin & Deno, 1993a, 1993b), written expression (Espin, De La Paz, Scierka, & Roelofs, in press; Espin et al., 2000; Parker, Tindal, & Hasbrouck, 1991a, 1991b; Tindal & Parker, 1989; Watkins & Lee, 1992), mathematics (Foegen, 1995), as well as on content-area learning in social studies (Espin, Busch, Shin, & Kruschwitz, 2001; Espin, Shin, & Busch, 2002) and science (Busch, Lembke, Seo, & Espin, 2002). In this article we review the research on the development of CBM measures in the content areas.
Curriculum-Based Measurement in the Content Areas

The initial research on the development of CBM measures in the content areas was conducted by Tindal, Nolet, and colleagues (for a review, see Espin & Tindal, 1998, and Tindal & Nolet, 1995), who focused on the development of “critical thinking” measures—measures that reflected the type of critical thinking skills required of students across content-area curricula. Later research by Espin and colleagues took a different approach, focusing on the identification of general outcome measures—measures that reflected students’ general performance in such content areas as social studies or science (for a review of the concept of “general outcome measurement,” see Fuchs and Deno, 1994). It is this later research that will be reviewed here.

Espin and colleagues examined three potential CBM measures in the content areas: reading aloud, maze selection, and vocabulary matching. Espin and Deno (1993) conducted the initial research on identification of general outcome indicators for secondary-level students in content areas. Because the number of words read aloud in one minute had correlated highly with other measures of reading performance at the elementary-school level (Deno & Fuchs, 1987), this study focused on the use of a read-aloud measure to predict performance on a study task in science and English. The purpose was to examine the relation between performance on a CBM reading-aloud task and a task requiring students to locate information in a text to answer comprehension questions. Participants were 10th-grade students who read aloud for one minute from three textbook passages in science and three in English. Following the reading-aloud procedure, students read through a longer passage and answered comprehension questions. After completing this study task, students again read for one minute from the three science and English textbook passages.

Results of the study revealed low/moderate correlations between the number of words read correctly and performance on the study tasks in both science and English ($r = .37$ for both areas). These correlations were substantially lower than those found between words read aloud and general reading performance at the elementary-school level, where correlations ranged from $.70$ to $.90$ (Deno, 1985; Fuchs, Fuchs, & Maxwell, 1988). The small magnitude of the correlations between the read-aloud measures and the study tasks prompted the search for other measures that might have greater predictive validity for performance in the content areas.

In a reanalysis of their data, Espin and Deno (1994-1995) examined the validity of a vocabulary-matching measure (originally used to assess student background knowledge) for predicting performance on the study task. Each vocabulary-matching probe consisted of 10 vocabulary terms and 12 definitions. The terms were selected from italicized/bolded words in the students’ texts or from words deemed specific to the content of the passage by the investigators. Definitions were formulated using The American Heritage Dictionary of the English Language (Davies, 1970). Participants were given 10 minutes to match the terms to their proper definitions.

Findings from this study revealed that the magnitude of the correlations between the vocabulary-matching measure and the study task in both science and English was slightly stronger ($r = .40$ and $r = .44$, respectively) than that for reading aloud. In a regression analysis, when vocabulary matching was entered first, it accounted for a significant proportion of the variance in the study task ($R^2 = .17$ and $.20$ in science and English, respectively); reading aloud did not add to the equation. Conversely, when reading aloud was entered first, vocabulary matching contributed a significant amount to the equation ($R^2 = .11$ and .09 in science and English, respectively). These results suggested that vocabulary matching might prove to be a good general outcome measure for content-area learning.
In a follow-up study, Espin and Foegen (1996) investigated the relative contributions of three potential CBM measures for predicting secondary students' performance on content specific material. The CBM measures were reading aloud from text, vocabulary matching, and maze selection. The criterion variables represented tasks commonly required of students in their content-area courses: comprehension (understanding what is read), acquisition (learning information following instruction), and retention (remembering information.) Participants in the study were 184 sixth-, seventh-, and eighth-grade students from a large urban middle school.

The findings of this second study replicated those of Espin and Deno (1994-1995). That is, all three CBM measures adequately predicted performance on the content-area tasks (with correlations ranging from .52 to .65.), but vocabulary matching was the best predictor of performance with correlations ranging from .62 to .65. In a series of regression analyses, vocabulary matching accounted for a substantial proportion of the variance in all the criterion tasks ($R^2 = .43, .41, \text{ and } .38$ for comprehension, acquisition, and retention, respectively), with no additional variance accounted for by maze selection or the reading aloud tasks. When maze selection or reading aloud were entered first, however, vocabulary matching contributed significantly to the equation ($R^2 = .05 \text{ to } .11 \text{ and } R^2 = .11 \text{ to } .18$ for maze selection and reading aloud, respectively).

Although the results of the Espin and Deno (1994-1995) and Espin and Foegen (1996) studies revealed the potential for vocabulary matching to be a valid indicator of performance and progress in the content areas, three questions remained. First, would vocabulary matching be a valid indicator of performance in an actual content-area classroom using classroom materials had not been examined. Second, what role did reading play in the prediction of student performance? In the research to that point, results of regression analyses had suggested that the strength of the vocabulary matching measure came from its combination of reading and content-knowledge requirements. That is, vocabulary matching added to predictive equations even after reading-aloud and maze were entered, but the reverse did not hold true. But was reading even necessary in the prediction of performance, or would a measure requiring content knowledge alone predict performance as well as, or better than, a measure requiring reading and content knowledge combined? Third, would vocabulary matching be a valid and reliable indicator of growth over time? In the research to that point, vocabulary matching had been examined as a static measure only. The validity of the measure for reflecting growth over time had not been examined. These three questions were addressed in two subsequent research studies.

**The Validity of Vocabulary Matching for Predicting Performance in a Content-Area Classroom**

To examine the validity and reliability of vocabulary matching in an actual classroom setting, Espin, Busch, Shin, and Kruschwitz (2001) and Espin et al. (2002) conducted two studies in a seventh-grade social studies classroom. In the first study (Espin et al., 2001), the reliability and validity of vocabulary matching for predicting performance in a social studies class was examined. To better assess the role of reading in the prediction of student performance, two types of CBM measures were examined: a vocabulary-matching measure in which the students had to read terms and definitions themselves (student-read probes) and a measure in which the examiner read the terms and definitions to the students (administrator-read probes).
Vocabulary-matching probes were 20 terms and 22 definitions, chosen from the classroom textbook, teacher's lecture notes and teacher-made tests. Criterion variables were a researcher-developed knowledge test, student grades, and performance on the social studies subtest of the Iowa Test of Basic Skills (ITBS) (Hoover, Hieronymus, Frisbie, & Dunbar, 1992).

Results revealed that both the administrator- and student-read probes were reliable predictors of student performance. Alternate-form reliability ranged from .58 to .87 for administrator-read measures and .63 to .81 for student-read measures. Combining scores across two adjacent measures increased reliability ($r = .70$ to $.85$ and $r = .76$ to $.88$ for administrator- and student-read measures, respectively). With one exception, validity coefficients ranged from $r = .64$ to $.81$ for student-read probes and from $r = .73$ to $.84$ for administrator-read probes. The exception was the low to moderate correlations found between the vocabulary matching measures and course grades ($r = .34$ and .51 for administrator- and student-read probes, respectively). The authors attributed these lower correlations to the multitude of factors other than learning that contributed to course grades (e.g., attendance, class participation, homework completion). Results of this first study suggested that vocabulary matching was a reliable and valid predictor of student performance in a content-area classroom, and that reading was not a factor in the measure's predictive power.

In a second study (Espin et al., 2002), student progress over time was monitored, as students completed one vocabulary-matching probe per week for 14 weeks. The reliability and validity of the vocabulary-matching measures for reflecting progress over time was examined. Specifically, the sensitivity of the measures for revealing longitudinal student growth and for discerning interindividual differences in growth rates was examined. Further, the validity of the growth rates for the vocabulary measures with respect to improvement on the criterion measures was also investigated.

Hierarchical linear modeling (HLM) was employed to investigate the reliability and validity of the vocabulary-matching measures as growth measures. Specifically, unconditional HLM models were used to examine the reliability of the probes for indexing student growth over time and for examining interindividual differences in student growth rates. The validity of the measures was investigated by statistically examining the relations between the growth rates on the vocabulary measures and the criterion measures — researcher-developed knowledge test, course grades, and scores on the social studies subtest of the ITBS.

Results of the HLM analyses revealed that the student- and administrator-read vocabulary measures were both sensitive to detecting improvements in students' performance over time. However, tests of the growth-parameter variance revealed that the administrator-read measure was not sensitive to revealing interindividual differences in growth rates. This lack of sensitivity precluded its use as a viable progress monitoring measure and excluded it from further analysis. The validity of the growth-rates estimated by the student-read vocabulary measure, in contrast, was significantly related to the criterion variables. That is, students who had higher scores on the knowledge test, higher course grades, and higher scores on the ITBS also had higher growth rates on the student-read vocabulary probes.

In sum, the findings of the second study revealed that only the student-read vocabulary measure was sensitive to indexing growth over time, implying that reading was an integral part of content-area learning and, as such, should be included as a part of a measure designed to index growth over time in content-area learning.

Efforts are currently underway to replicate the social studies research in science. Initial
analyses reveal a similar pattern of results in terms of the use of a student-read vocabulary-matching probe as a measure of performance and progress in the content areas (Busch et al., 2002)

Identification of a technically adequate measure for monitoring content-area learning is an important first step in helping special educators better meet the needs of their secondary-level students. Such a measure must provide information that aids teachers in making instructional decisions about their students. In the following section, we provide an example of how the measure could be used in a school setting to make instructional placement decisions for students with learning disabilities.

Use of Curriculum-Based Measurement in the Content Areas

Mr. Lanigan is a special-education teacher at Franklin Middle School. He has two eighth-grade students, Meaghan and Patrick, on his caseload, who are placed in a general education world geography course. In order for Meaghan and Patrick to be successful in the geography class, Mr. Lanigan spends considerable amounts of time modifying assignments and providing accommodations on classroom quizzes and tests. Although both students are receiving passing grades in the class, Mr. Lanigan is concerned that they may not be learning the class curriculum, and may be earning passing grades only because they turn their work in on time and participate in class. Mr. Lanigan decides to use the vocabulary-matching measure to monitor the progress of both students and determine whether the world geography class is an appropriate instructional setting for the students.

Mr. Lanigan begins by generating vocabulary probes from the glossary of the textbook used in the world geography class, randomly selecting 20 terms and 22 definitions (including two distractors) for each probe. He places all the terms in alphabetical order down the left side of a piece of paper and randomly places all of the definitions down the right-hand side. To assess Meaghan and Patrick's current levels of performance, Mr. Lanigan gives each student four vocabulary probes over the course of two weeks. For each probe, the students have 5 minutes to match terms to definitions. Mr. Lanigan graphs the data for each student and sets a long-range goal based on an increase of one correct match every two weeks.

Once a week, Mr. Lanigan collects data on Meaghan and Patrick and graphs the mean score of each pair of adjacent probes. Using these mean scores Mr. Lanigan compares the actual progress of Meaghan and Patrick with their long-range goal (see Figures 1 and 2). Visual inspection of Meaghan's graph reveals that she is progressing on the vocabulary probes at a pace equal to that set by Mr. Lanigan as a long-range goal. Meaghan's growth rate indicates that the instruction in the world geography class is effective for her and that, therefore, her placement in the class is appropriate.

Inspection of Patrick's graph, however, reveals that he is showing little growth on the vocabulary probes, and has fallen well below the long-range goal set for him. His growth rate indicates that Patrick is not learning the course content in the world geography class, and that a change in placement may be in order. After meeting with the world geography teacher and Patrick's parents, Mr. Lanigan decides to change Patrick's placement to a world geography class co-taught by Mr. Lanigan and a general education social studies teacher. In this class, the curriculum can be better tailored to meet Patrick's academic needs. After the placement change, Mr. Lanigan continues to use the vocabulary probes to monitor the effectiveness of the instruction being provided to Patrick in his new placement, and also continues to monitor Meaghan's progress in her world geography class.
Conclusion

Research on the use of curriculum-based measurement at the secondary level has focused on reading, written expression, and, most recently, content-area learning. In the content areas, vocabulary matching has proven to be a valid and reliable indicator of performance and progress in the content areas. Theoretically, the validity of the vocabulary-matching measure can be explained in terms of the skills required to complete the task. Both the vocabulary-matching measure and content-area learning require that students be able to read content-area text fluently, know the meaning of vocabulary terms, and hold content-area information in working memory. As students learn more in their content-area classes, it is likely they will be able read text more fluently, define more terms, and hold information in working memory more easily. These skills will also enable them to do better on the vocabulary-matching measures.

Practically speaking, the studies reviewed in this article suggest that vocabulary matching can be used by teachers to predict students’ performance in content-area classes and monitor students’ learning in those classes.

Although the results related to the use of the vocabulary-matching probes are promising, two cautionary notes are in order. First, the alternate-form reliability of the measures is somewhat lower than desired. Combining two measures can increase reliability; however, this is burdensome for teachers. Future research should examine methods for increasing the reliability of the measures. Second, the research thus far has focused only on the development of technically adequate mea-
sures. The next and most important step is to examine the achievement effects associated with the use of a progress measurement system. In other words, when teachers use CBM to monitor student progress and evaluate the effects of instruction, does student achievement improve?

In conclusion, vocabulary matching has been shown to be a valid and reliable indicator of performance and progress in the content areas. Teachers can use vocabulary matching to identify students at risk for failure early in the school year, and those students can then be monitored to examine the extent to which they are learning content-area material. Thus, the data can be used to evaluate the effects of instruction, and ultimately of a placement, for a student. The effects of this early identification and monitoring on overall student achievement has yet to be examined.

References


