CBM progress monitoring in reading and foreign-language learning for secondary-school students

Chung, S.

Chapter 6

Summary and general discussion

Introduction

Curriculum-Based Measurement (CBM) is a progress-monitoring system designed to be used by educators to screen performance, to monitor growth, and to evaluate the effectiveness of interventions on the growth of students with learning difficulties (Deno, 1985, 2003). Implementation of CBM for monitoring growth and evaluating instruction involves frequent measurement of student performance. Scores from the measures are placed on graphs that depict growth over time and provide information to educators about how effective instruction has been.

Although there has been a large body of research conducted on CBM (Espin et al., in press, or see also Chapter 2; Foegen et al., 2007; McMaster & Espin, 2007; Wayman et al., 2007), there are notable gaps. First and foremost, relatively few studies have been conducted at the secondary-school level, and of those, few have focused on reading, and none on foreign-language learning. Second, relatively few studies have been conducted at what L. S. Fuchs (2004) refers to as Stage 2 or Stage 3 CBM research. Stage 2 studies focus on the technical adequacy of scores from the measures as indicators of growth, and Stage 3 on the instructional utility of CBM, that is, on the effects of implementing CBM on teacher instruction and student achievement. The majority of CBM studies have been Stage 1 studies, which focus on the technical features of static (performance) scores at one point in time. The disproportionate focus on Stage 1 may be due in part to the popularity of using CBM measures for screening, or to the fact that studies at Stage 1 are less costly and time-intensive than studies at Stages 2 or 3 (L. S. Fuchs, 2004; Wayman et al., 2007). In addition, Stage 1 studies are often the first step in CBM research: promising measures are identified in Stage 1 studies, which then can be more closely examined in the more costly and time-intensive Stage 2 and 3 studies.

The studies in this dissertation addressed gaps in CBM research. First, the research focused on the development of CBM progress-monitoring measures at the secondary-school level in reading and foreign-language learning. Second, the studies in reading were Stage 2 studies, and focused on the technical features of the slope/growth rates produced by reading CBM measures.

The essential characteristics of CBM measures as outlined by Deno (1985) served as a framework for this research project. First, the measures had to be simple, efficient, and inexpensive so they could be used by educators to frequently and repeatedly (for example, weekly) monitor student growth. Second, scores from the measures had to be easy to understand so that the results could be clearly and correctly communicated to parents, teachers or students. Finally, and most importantly, scores from the measures had to be reliable and valid *indicators* of performance level and growth in an academic area so that improvement in scores reflected general improvement in the academic area. In this way, the measures could be used by educators to monitor the general growth of students in an

academic area (for example, reading), and to evaluate the effects of their instruction on that growth. In our studies we wanted to obtain a broad view of the validity of CBM scores by adopting Messick's (1989) recommendation to build evidence for validity across various criterion variables. Thus, we included multiple criterion variables that were assumed to represent performance level or growth in the academic area of interest and examined the pattern of relations of the criterion measures with the CBM scores.

In the following sections, the results of the studies included in this dissertation are briefly summarized and reviewed. This is followed by a general discussion of the implications of the findings.

Development of CBM Measures in Foreign-Language Learning

To our knowledge, Chapter 3 on the development of CBM measures in foreign-language learning was the first CBM study to be published in this area. The study was Stage 1 CBM research (L. S. Fuchs, 2004) and focused on the technical adequacy of CBM scores as indicators of (static) performance in foreign-language learning. Specifically, we examined the reliability and validity of CBM maze and Dutch-to-English and English-to-Dutch word-translation scores for 8th- and 9th-grade students. Stage 1 research was appropriate given the fact that there has been little previous CBM research in the area of foreign-language learning and with students in secondary school. The results of this initial study can be used to guide future Stage 2 and 3 studies.

In general, scores on 2-min maze and 2-min word translation (either Dutch-to-English or English-to-Dutch) were reliable and valid indicators of performance level in foreign-language learning for secondary-school students. Different administration times and scoring procedures were examined. The results of the reliability analyses revealed that scores on all three measures that were administered for two minutes were more reliable than for one minute administration times. With regard to the scoring procedure, the correct-minus-incorrect maze scores were slightly more reliable than the number of correct maze scores, whereas for both word-translation tasks the number of correct translations were more reliable than the correct-minus-incorrect translations. The use of a guessing rule for the maze or a spelling rule for word translation did not affect the reliability of the scores on the measures.

Based on the results of the reliability analysis, and on factors such as ease, efficiency of scoring, previous research, and so forth, a subset of scores were selected for subsequent validity analysis. In general, the pattern of results supported the validity of scores from 2-min maze and 2-min word-translation tasks (either English-to-Dutch or Dutch-to-English) as indicators of general foreign-language performance level. No clear pattern of differences emerged across scores from the three potential measures; however, validity results were generally better for students in their 3rd year of English than for students in their 2nd year of English. Finally, we also explored the effects of combining scores across measures for

students in their 3rd year of English (9th grade). A combination of scores from maze and the English-to-Dutch word translation provided the best prediction of English course grades, suggesting that, if the measures are to be used for screening purposes, the use of scores from two CBM measures might be better than scores from one measure alone. Including the scores from the Dutch-to-English word translation did not improve prediction.

The extent to which scores from these measures can be used to monitor student growth and inform teachers' decision-making needs to be investigated in future Stage 2 and 3 studies.

Development of CBM Measures in Reading

Two studies were conducted to examine CBM measures in reading for secondary-school students. The studies were Stage 1 and 2 CBM research (L. S. Fuchs, 2004) and focused on the technical adequacy of CBM maze scores as indicators of performance level and growth in reading. Specifically, the alternate-form reliability, sensitivity, validity, and stability of scores or change in scores from CBM maze were examined for 7th-grade students. Both linear and nonlinear (logistic) growth trajectories were examined.

In the first study that we conducted on the development of CBM measures in reading, scores on 2-min mazes were found to be reliable, sensitive, and valid indicators of performance level and growth in reading for secondary-school students. Results for alternate-form reliability analyses revealed that the reliability was moderate-good. Sensitivity analyses showed that maze scores were sensitive to change in performance, and students differed in growth trajectories. This was the case for both linear and nonlinear growth, but nonlinear growth fit the maze data better. The overall pattern of results showed that the CBM maze task was valid and supported the use of CBM maze scores as indicators of performance level and growth in reading. In general, students with higher reading proficiency, as assumed by their dyslexia status, school level, and scores and change in scores on a reading test, had higher scores and grew steeper on the maze than students with lower reading proficiency. The pattern of results were similar for linear and nonlinear growth.

The findings from the study provided tentative support for the technical adequacy of scores from CBM maze measures as indicators of reading performance level and growth for secondary-school students; however, the results were based on 23 weeks of progress monitoring. If educators are to use progress data to inform instruction, then they must be able to make decisions in a timely fashion, for example, after 8 or 10 weeks. Therefore, the next study investigated the stability of the maze growth rates or slopes under various conditions. This was the first study to examine the stability of maze growth, and the first to examine the stability produced from CBM reading measures at the secondary-school level.

Results of this subsequent study revealed that the stability of maze slopes was dependent on a number of factors, including the duration of the progress-monitoring

period, the data-collection schedule (i.e., how many data points and often data were collected), and the variation in individual maze scores (reflected in the standard error of estimate). Longer durations, denser schedules, and lower variation in maze scores increased the stability of the maze slopes. However, relatively large instability was seen in the first few weeks of data collection, regardless of the schedule or maze variation, but decreased with longer duration. The implications are that educators should monitor students' growth for at least 10 to 12 weeks – assuming that mazes were administered once per week under 'typical' conditions – before the maze slope would stabilize. Finally, the effects of assuming nonlinear growth on maze variation were examined, but did not result in reduced variation in scores.

The Use of CBM Scores and Future Directions

The results of our studies contributed to the development of CBM progress-monitoring measures at the secondary-school level, and were Stage 1 CBM research (L. S. Fuchs, 2004) in foreign-language learning, and Stage 1 and 2 CBM research (L. S. Fuchs, 2004) in reading. These stages focused on whether CBM scores were suitable for monitoring learning progress. In this section, the results are compared with previous research and the use of the CBM scores are discussed separately for foreign-language learning and reading.

Foreign-language learning

It is difficult to compare the results of our study on the development of CBM measures for foreign-language learning to previous studies because there have been few (if any) such studies. However, it is possible to compare our results to CBM secondary-school studies in related areas, for instance, studies on the development of measures for English-language learners (ELLs) in the United states (McMaster et al., 2006), studies that have examined maze scores as indicators of reading performance level (Espin et al., 2010; Tichá et al., 2009), or studies that have examined scores from a vocabulary-matching task as indicators of content-area learning (for example, Busch & Espin, 2003; Espin et al., 2001; Espin et al., 2013; Espin & Deno, 1994-95). Such studies provide a context within which to evaluate the results of our study.

The reliability results of the foreign-language measures could be considered good, given that the reliability is either similar or somewhat better than what has been found in previous studies (Espin et al., in press, or see also Chapter 2; McMaster et al., 2006). However, the validity results were somewhat weaker than those found in previous studies (Espin et al., in press, or see also Chapter 2). It is worthwhile to note that the 'weaker' relations were mostly seen for 8th-grade students, whereas the validity results for 9th-grade students were similar to those of previous studies (Espin et al., in press, or see also Chapter 2; McMaster et al., 2006). It is possible that students need to reach a basic level of general language-proficiency before the CBM scores can predict performance in foreign-language learning adequately, and that the 8th-grade students who were in their second year of English

instruction, had not yet reached this basic level. Perhaps 9th-grade students had reached this level because they had received an extra year of English instruction, and for instance the English language learners in the previous studies had also reached this basic level because they lived in a 'rich' environment where English was spoken regularly. Nevertheless, it is too soon to draw firm conclusions, especially given that the validity analyses in the foreign-language learning study were conducted within grade level and school- or reading test level, and sample sizes were therefore small.

An interesting result from the foreign-language study was that a combination of scores from two CBM measures, maze and English-to-Dutch word translation, predicted performance in foreign-language learning better than a single measure alone. Although this result should be interpreted carefully due to small sample sizes, this finding was similar to that of Espin and Foegen (1996), in which a combination of scores from maze and vocabulary-matching predicted the performance of content-area tasks better than scores from only one of the measures. It might be worthwhile to examine such predictions on performance in future research.

The results of the study on CBM foreign-language learning are promising, but it is important that such results be replicated in future research, if possible with a larger sample size and wider range of grade levels. Most importantly, additional research should be conducted to examine the CBM scores as indicators of growth in foreign-language learning at secondary-school level, and to examine the effects of teachers' use of such progress-monitoring measures on their instruction and on student achievement.

Reading

Although the technical adequacy of scores from CBM measures in reading at secondary-school level has been examined more extensively than for foreign-language learning, the research has still been sparse, especially compared to the amount of research in CBM reading that has been conducted at elementary-school level (see Wayman et al., 2007). Thus, in our study, we focused on the alternate-form reliability, sensitivity, validity, and stability of maze scores as indicators of reading performance level and growth at the secondary-school level.

The reliability results of the 2-min maze scores were slightly weaker than the 3-min maze scores used in studies that also included frequent (i.e., weekly) measurement of CBM measures (Espin et al., 2010; Tichá et al., 2009). Such results were acceptable given that the reliability in general was moderate-good. Nonetheless, a lower reliability may affect the variation in maze scores, and in turn, the stability of maze slopes of individual students. The sensitivity and validity results were similar to what was found in most previous studies (Espin et al., 2010; Ticha et al., 2009; Tolar et al., 2014; Tolar et al., 2012). The replication of these results was important and necessary because previous studies included only small sample sizes (Espin et al., 2010; Tichá et al., 2009), or included only five measurements that were

too few for making timely instructional decisions (Tolar et al., 2014; Tolar et al., 2012), whereas in our study, progress monitoring occurred frequently (i.e., weekly) for 23 weeks for a large sample of students. Finally, since our study was the first to examine the stability of scores from CBM reading measures at the secondary-school level, and the first to examine the stability of scores produced by the maze measure, we compared our results to studies at elementary-school level focused on CBM reading-aloud scores. The results on the stability of maze slopes were similar to the findings for the stability of reading-aloud slopes at the elementary-school level (Christ, 2006; Christ et al., 2013).

It is positive that our results confirmed earlier findings. However, although the results were consistent with previous studies, it is important to note that the maze growth rates differed greatly across studies. For instance, growth rates in the current study were similar to the growth rates found in Tolar et al. (2012), but more than five times smaller than growth rates found in other studies (Espin et al., 2010; Tichá et al., 2009). The various possible explanations for these differences were discussed in detail in Chapter 4; however, we emphasize here that the differences are of concern, especially in relation to the stability of the growth rates. If expected growth rates are smaller than the amount of instability in individual maze slopes, students' reading growth might not be adequately represented, and this might lead to incorrect interpretations and instructional decisions. A more detailed discussion of this issue and possibilities to reduce maze score variation in order to improve the stability of maze slopes are provided in Chapter 5. It is important that in future research normative growth rates are established by administering a validated maze passage set to a representative sample of students. In addition, future research should also focus on reducing maze score variation around the slope. Such research would provide educators with information regarding what they should expect from their students, and thus would help with setting long range goals in learning when using CBM.

Although there is much that remains to be examined in future research, we did find that, in general, scores or change in scores on CBM maze were reliable, sensitive, and valid indicators of reading performance level and growth at secondary-school level. The amount of stability given the expected growth rates is still a matter of concern, which needs to be addressed in future research.

One final related issue is the functional form of the expected growth of students. Although we found that nonlinear growth better represented the data, it appears that the difference is small given that the variation in maze slopes did not differ when either linear or nonlinear was assumed. It would be worthwhile to examine this issue in future research.

Conclusion and Final Remarks

This dissertation project contributed to the knowledge base of Stage 1 and 2 CBM research (L. S. Fuchs, 2004) at secondary-school level, namely to the suitability of using scores from CBM measures as indicators of performance levels and/or growth in foreign-language

learning and reading. Scores from CBM maze and word translation emerged as potential measures for CBM progress-monitoring in foreign-language learning. Moreover, the results in reading for CBM maze measures strengthens earlier research, and provides additional evidence for the use of maze scores in CBM progress-monitoring at secondary-school level.

Based on our results and given the intended use of the scores, we carefully recommend that 2-min CBM maze or CBM word-translation scores could be used as indicators of foreign-language learning performance levels, although more research is necessary, especially research focused on growth. For monitoring reading growth, we recommend the use of 2-min CBM maze scores for 7th-grade students, but, the measure or administration conditions must change slightly in order to improve the stability of the maze slopes. Moreover, we currently recommend that CBM reading growth can be presented as a linear rather than nonlinear growth trajectory, because from our research it could not be concluded that one approach was *much* better than the other, and linear growth trajectories provide higher practical advantages and possibly higher expectations from educators for their students' performance (see also Chapter 4).

As a last point, we want to emphasize that the 'real test' for CBM progress monitoring lies at Stage 3 research. It is one thing to develop a measurement system that has evidence of reliability and validity for monitoring the growth of students in an academic area, but quite another to determine whether implementation of such a system leads to improvements in the instruction offered to students who struggle, and ultimately to improvements in their achievement. The ultimate test of CBM is to examine whether teachers' use of CBM measures for monitoring the growth of secondary-school students who struggle in reading and/or foreign-language learning leads to improvement in teachers' instruction and to improvements in student performance. A related question is whether implementation of CBM progress monitoring leads to changes in teachers' beliefs about the extent to which they can influence the performance and growth of students who struggle.

In sum, the ultimate goal of CBM implementation is to provide teachers a method that allows them to 'scientifically' examine and evaluate the effectiveness of the instruction they deliver to students who struggle, and to create in teachers a sense of responsibility and a sense of urgency with regard to performance and growth for secondary-school students who struggle. Relatively speaking, educators have little time to influence the learning of students who struggle. Time cannot be wasted on the implementation of ineffective instruction for these students. In addition, educators must strive to develop in such students the maximal levels of skills possible in order to improve the likelihood of success for these students in their post-secondary school years. Perhaps CBM implementation can be a part of the development of effective and powerful instructional programs for students who struggle.