W. Corry Larson
Camilla Ward

Developing a Predictor of University Students' Course Grades Using Curriculum-Based Measurement: An Initial Investigation

University students (n = 69) enrolled in an introductory psychology course at an historically Black university completed nine curriculum-based measurement probes. Results for each student were graphed and trendlines were calculated. Trendline direction was examined to determine whether a possible formative assessment technique might be developed for use at the university level. Suggestions for further research concluded the investigation.

Assessment is an integral part of effective education. Without valid, reliable and efficient assessment, teachers have no idea whether students have learned what they have been taught. Research findings provide support for Curriculum-Based Measurement as effective and efficient classroom assessment (e.g., Fuchs, Fuchs, & Hamlett, 1994; Gable, Arllen, Evans, & Whinnery, 1997).

McDaniel, Evans-Hampton, Skinner, Henington, and Sims (2002) described the characteristics of curriculum-based measurement (CBM) as procedures that are

> ...useful in making educational decisions for specific students. Specifically, CBM procedures are brief; educators can construct multiple parallel forms from the students' curricula; and results yield rate data that are extremely sensitive to changes in academic skills.... Thus, CBM data can be used to evaluate students' learning rates frequently and over brief periods of time.... These characteristics allow educators to use CBM to compare more than one instructional intervention to determine which intervention maximizes learning rates for individuals. (p. 1)

A large body of research (Deno, 1985; Fuchs, Deno, & Mirkin, 1984; Fuchs & Fuchs, 1986; Fuchs & Fuchs, 1988; Kim, 1998) has provided support for using CBM, a set of standard, simple, objectively-scored, short-duration fluency measures, or "probes" to measure student achievement in important areas of basic skills or literacy, such as reading, spelling, written expression, and mathematics. The data from these measures are designed to frequently monitor students' growth in important skills domain relevant to school outcomes (Shinn & Baimonto, 1998), and, when graphed, provide an easy-to-interpret means of determining student progress (Council for Exceptional Children, 2005). These probes are superior to teacher-made quizzes because they can be constructed with adequate reliability, and therefore consistently compare the same construct across time and instruction. Teacher-made quizzes may not be constructed to measure reliably, and may therefore not yield data that can be compared over time (Deno et al., 1982).

An equally large body of research has established support for the technical adequacy, validity, standardization, and reliability of CBM to formatively assess student progress (Bean & Lane, 1990; Deno et al., 1982; Espin & Deno, 1993). This technical adequacy is useful not only to researchers, but also to teachers who wish to have an indication during the class that their students are on the path toward mastering the material being taught.

In CBM, academic performance is sampled through the use of dynamic direct observation procedures. All CBM scores are obtained by counting the number of correct and incorrect responses made in a fixed time (Deno, 2003). Scores, then, provide a measure of the level of fluency on a student's skill level.

One of the most distinctive and important features of CBM is that performance is reliably sampled repeatedly across time. As a dynamic indicator of student learning, CBM is sensitive to differences in student learning. According to Shinn and Baimonto (1998),

> Commercially available norm-referenced achievement tests are constructed to be sensitive to differences *among* individuals.... CBM, like commercially available tests, is also sensitive to differences among individuals. However, unlike commercial norm-referenced achievement tests, CBM is dynamic (i.e., sensitive) to differences *within* a person over time. They are designed to be sensitive to the short-term effects (i.e., 4 – 6 weeks) of instruction. When a student's skills change, the measures will detect this growth.... This sensitivity allows for a student's learning to be detected and documented. (p. 6)

CBM is a *formative* evaluation tool. Instead of evaluating a student's instructional program at the end of an instructional period (i.e., summative evaluation), when it is too late to change an on-going program, CBM is used to assess a student's progress continuously during instruction (formative evaluation). CBM allows the teacher to make regular—and repeated—decisions about whether student progress is satisfactory or unsatisfactory. If CBM shows that the student is progressing satisfactorily, the program is maintained. Should CBM indicate that the student is not making satisfactory progress, the instructional program may be changed in some meaningful way with the goal of improving student outcomes (Shinn & Baimonto, 1998).

Research on CBM in the assessment of basic skills has been conducted for the most part on students in the elementary and middle school grades (Fuchs, Fuchs, & Maxwell, 1988). Students in both special education (Frank & Gerkin, 1990) and general education (Baker & Good, 1995) have participated in this research. Results from research in which CBM was employed suggested that because CBM uses stimulus materials drawn from a student's local curriculum, students from minority backgrounds may be more adequately assessed by CBM than by achievement or intelligence tests that use a broad sample (usually national) of students to determine norming distribution (Shinn, Collins, & Gallagher, 1998). Deno (1985) suggested that because a problem exists when a student does not perform the academic behavior(s) expected in a particular curriculum, such problems are seen as situational in CBM, rather than inherent in a particular student.

The majority of CBM research at the secondary level has been conducted in the area of content-area learning, with a small body of work conducted in written expression and mathematics (Shinn & Baimonto, 1998). Deno (2003) reported research findings related to the use of measures for determining general reading proficiency at the secondary level, as well as research on measures focused solely on content-area reading.

Allinder (1995), Bean and Lane (1990), and Sticht (1999) reported research conducted using CBM in adult education, but apparently no studies have been conducted using CBM in the university classroom. An extensive search of the literature, using both publicly available search engines (Google and Google Scholar; Search ERIC.org), and for-pay Internet libraries (Highbeam Research), yielded no research articles on this subject.

In line with the research available on using CBM with adult learners, a short review of the use of CBM with secondary (specifically high school) students would seem to offer at least a partial background for studying CBM measures for university students. This use of the CBM research should, however, be used with caution because the two populations may be sufficiently different so that what is known about secondary students may not hold true with university students.

Espin and Tindal (1998) reported on three CBM measures (read aloud from text, a cloze-like procedure called maze, and vocabulary matching) that have been developed to assess secondary-level students' progress in general reading skills and content-area reading skills. All three of these CBM measures proved to be adequate predictors of student performance, but Espin and Deno (1994-1995) and Espin and Foegen (1996) found that, of the three, the vocabulary matching task was a better measure of performance on reading and content-area learning tasks than are the other two. As Espin and Tindal (1998) suggested, "As students increase their knowledge in a content area… we could imagine that they would also increase their knowledge of words specific to that area" (p. 231). The present investigation was intended to further explore the relation of monitoring student progress using CBM to content learning.

The investigators of this study considered three questions:

1. Did reading skill, as measured by the Accuplacer (College Entrance Examination Board, 2002) reading test, vary with whether there was a positive relationship (ascending trendline) or negative relationship (descending trendline) in a participant's graph of the results of CBM probes?
2. Was there any relationship between the direction of the trendline of the participant's graph of CBM probe results and the grade the participant earned in the class?
3. Over the course of the class, was there any trend in the number of ascending, descending, or horizontal trendlines across the class?

Number 3 relates to the second part of Espin and Tindal's (1998) and Shinn's (1998) observation that students who experience difficulty in content-area reading may be placed into two groups: those with difficulties due to general reading deficits and those whose difficulties were due to reading deficits specific to the content area. A preponderance of ascending trendlines in student graphs over the duration of the course might indicate that class participants were increasing their reading skills in areas specific to the course content. That is, in terms of the Espin and Tindal (1998) and Shinn (1998) observation, over the course of the class, ascending graph trendlines might indicate that students were gaining a better understanding of the course material, thereby improving their reading skill in this specific content area. Whether a student's trendline was ascending, descending, or remaining horizontal might be an indication of that student's success in learning the class material. If validated, this tool would give an instructor an easy-to-use, straightforward, but reliable and valid, instrument for formatively ascertaining an individual student's progress throughout the class. Descending or horizontal trendlines would be an "early warning system" of student difficulty in the class. This "early warning" would then prompt additional instructor intervention before the student had lost out on so much understanding of the class material that passing the class would become problematic.

Method

Participants

Sixty-nine university undergraduate students (22 men and 47 women, mean age = 19.7, *SD* = 2.3), who were enrolled in two introductory psychology courses at an historically Black college located in the Mid-Atlantic area of the East Coast, agreed to be participants in this project. Prior to collecting data, approval was obtained from the University's Institutional Review Board for the project, and all participants signed informed consent forms.

Demographic backgrounds

Participants self-reported their racial category as African American (72%), White (9%), and Other (1%), Other plus African American (7%). One participant refused to classify him/herself. The combined mean overall GPA for all participants was 2.7, on a grade scale of 4 (A) to 1 (F). The participant group was composed of fifty freshmen (72%), 13 sophomores (19%), six (9%) juniors, and one (1%) senior. (Totals do not add up to 100% because of rounding.)

Procedure

The investigators used the results of a reading test required of all students entering the University to estimate general reading ability of the participants. The investigators then constructed and administered probes to the participants. Finally, the investigators recorded and processed the resulting data. Participants were not paid for their participation but were promised an addition of five extra points to their class grade for every vocabulary probe they completed. A student's overall class grade was calculated by determining the ratio of the total points the student earned on exams, papers, and other assignments, to the total number of possible points. The maximum of 45 points a student could earn by completing all nine probes was in addition to all other points possible. Since the total number of points offered in the class was 650, the most that these "extra" points could account for was 7% of a student's grade, assuming that the student earned all other available points.

Reading ability. As a means of estimating participants' reading ability, Accuplacer (College Entrance Examination Board, 2002) reading scores and percentile ranks were obtained for all participants. Upon entrance to the university, all students (except for transfer students) were required to take Accuplacer placement tests to determine whether or not the student needed to be enrolled in remedial classes. The Accuplacer tests had a total possible score of 120, and had been used in a variety of higher-education settings throughout the country. The test also provided percentile ranks, which were determined to be of more immediate interest than the scores, because the percentiles represent a national sampling and, therefore, allow for comparison with other users of the test. Table 1 is provided to illustrate results from the reading subtest of the Accuplacer for the participants in this study.

Table 1
Means and Standard Deviations of Accuplacer Reading Scores and Percentile Rank for 69 Students

	Reading Scores	Reading Percentiles
Means	69	39
Standard Deviations	18	5

Since the investigators desired to aggregate the scores of the two classes in a effort to strengthen any findings that might result from this study, Accuplacer scores for the two classes were compared. The results of this comparison are provided in Table 2.

Table 2
Means and Standard Deviations of Accuplacer Reading Scores and Percentile Ranks for Class 1 (38 Students)

	Reading Scores	Reading Percentiles
Means	67	44
Standard Deviations	13	13

Means and Standard Deviations of Accuplacer Reading Scores and Percentile Ranks for Class 2 (39 Students)

	Reading Scores	Reading Percentiles
Means	72	35
Standard Deviations	16	16

Since the reading scores for Class 1 was within 10% of Class 2, the investigators felt that the two classes were sufficiently close in measured reading skill to be compared. Scores for both classes were therefore aggregated.

Probes. Generally following Espin and Tindal (1998, p. 226), a set of nine vocabulary probes was created by first scanning each page of the glossary of the textbook used in the psychology class (Myers, 2004) into a computer file. Each glossary page was represented by a separate file. Next, each file was assigned a number starting with the number one. Then each vocabulary word (along with its definition) on each page was assigned a sequential number, also starting with the number one. Finally, using a random number generator (Naace, n.d.), three sets of random numbers were generated. The first set was used to determine the page of the textbook glossary from which a word would be selected. The second set was used to determine which word and definition from that glossary page would be selected. The vocabulary words were placed on the probe in the order in which they were selected. The third set was used to determine the place on the probe page in which a definition would fall. Each set of three numbers was placed in a horizontal row on a three-column matrix. Each matrix had 21 rows, one for each word and its definition.

The probes were constructed by selecting the first number in the first row and first column of the matrix. That number was used to identify a computer file that represented a page from the textbook glossary. The second number in the same row was used to determine which word on the glossary page would be used first in the probe. Each glossary page had two columns. Starting at the top of the left column, words were counted down the page until the number represented by the second number in that same row was reached. Rules were developed to account for those occasions when the same word was selected for more than one probe. Finally, the third number in the same row of the matrix was used to determine the order of the selected word's definition on the probe page. The second word in the probe was placed below the first, and was chosen and placed in the same manner by selecting the first number in the second row of the matrix and using that to identify from which glossary page that second word would be chosen. The third number in the second row was used to determine the sequence of the second word's definition on the probe. Each definition was numbered, with the top definition assigned the number one, the second the number two, and so on, until there were 20 words and 21 definitions selected and placed on the probe. The definitions were numbered one to 21. One more definition than the number of vocabulary words was included to discourage identifying definitions by a process of elimination. The result was two columns, on the left the one for words and on the right the other for definitions. To the immediate left of each word was an underlined space. To complete the probe, the participant would place the number of the definition he or she chose on that underlined space so that the vocabulary word was matched with a definition. A space for the participant's name was provided at the top of the probe, as well as a space for the date the probe was completed. A cover sheet was attached as a first page for each probe to allow the experimenter to control when the participant first saw the probe. Nine probes were constructed in this manner.

Probe administration. Probes were administered in class once per week, for a total of nine probes for the entire semester. At the beginning of the semester, the probe administrator described what was asked of the participants, saying that 1) the participants were asked to identify as many of the listed vocabulary words as possible by placing the number of the correct definition next to the number of the word; 2) the participants had five minutes in which to complete this identification; 3) participants were asked to do their best and not to worry if they did not complete the matching of all vocabulary words to the definitions; and 4) participants were reminded that they would receive five extra points on their final grade for each probe they completed. As the semester continued and participants became more aware of the expectations, only a short summary of these expectations was given at each probe administration to remind participants.

The procedure for administering a probe was as follows: Graphs (which had been prepared beforehand) for each participant were distributed. Each participant had a personal graph so that the participant could chart his/her progress across probes. This participant graph was identical to the graph that the probe administrator kept on each participant to keep track of the number of vocabulary words correctly identified in each probe. Each participant graph contained the participant's name, a small vertical matrix to match probe scores to the date that a probe was given, and a graph. The vertical axis of the graph was labeled "Items Correct," and the horizontal axis was labeled "Probe Number." Probes were then passed out to all participants in the class, with the request from the probe administrator for participants not open their probes until told to do so. When all was ready, the probe administrator announced, "Please begin," while simultaneously starting a small countdown timer set for five minutes. Participants worked on their probes until the countdown timer's alarm sounded. At this point, the probe administrator asked participants to stop, close their probes, and exchange completed probes with another participant. The probe administrator unveiled a chart showing which vocabulary word number corresponded to which vocabulary number. Participants were asked to calculate the total number of correctly matched vocabulary words and definitions for the original participant's probe, place this number on the space on the probe cover sheet provided, and pass the probe back to its owner. The probe's owner then would note the score earned on the probe, and place that number in the appropriate slot on the small vertical matrix that listed the appropriate date for that probe. The probe owner then would place the probe score as a dot on his/her graph and connect that new dot to a previous dot on the graph. As time progressed, participants could see the progress they made by determining whether the direction of the graph's line ascended or descended. The probe administrator then would collect the completed probes and participant graphs, and the class would resume.

Data recording and processing. To record participants' scores, data were transferred from the probe sheets to a graph identical to the participants', but under the control of the probe administrator. The probe administrator checked each participant's probe to determine whether the data had been correctly collected and recorded. Data points were then recorded on the small vertical matrix of each participant's probe administrator graph in the same way that participants recorded their data, and the data point was transferred to the appropriate participant's administrator's graph.

At the end of the data-gathering period, data for each participant was transferred from the paper graphs maintained by the probe administrator to individual graphs that were created using Microsoft Office Excel 2004, v. 11.2.5 spreadsheets. A regression trendline was calculated for each graph. Then the participant's name, Accuplacer reading score and percentile rank, the participant's class (freshman, sophomore, etc.), and the grade the participant earned in the class were added to the graph.

Results

Reading scores vs. graph trendlines. Visual analysis of the participants' graphs was conducted by noting the direction of each participant's trendline: ascending, descending, or horizontal. For the use of visual analysis in the interpretation of graph trendlines, see Horner et al. (2005) and Parsonson and Baer (1992).

To investigate the first relationship (i.e., Did reading skill, as measured by the Accuplacer reading test, correlate with graph trendline direction?), a correlation between Accuplacer reading scores and graph trendline direction (1 = up, 2 = down, 3 = horizontal) using CORREL from the statistical function category of Microsoft Excel 2004 v.11.2.5 was performed. The result was a correlation of 0.05 between Accuplacer reading scores and graph trendline direction. The conclusion that the investigators drew from this data was that there appeared to be little relationship between direction of trendline (either ascending or descending) and measured reading skills for these participants.

In addition, an analysis was made seeking to establish whether a correlation existed between Accuplacer reading scores and the students' graph slopes. Using the SLOPE statistical function of Microsoft Excel, the slope of each student's probe number versus probe score was calculated. Using the CORREL statistical function of Microsoft Excel 2004 v.11.2.5, a correlation between Accuplacer reading scores and graph slopes was performed, resulting in a correlation of -0.13. This appeared to confirm the lack of relationship between Accuplacer reading scores and graph trendlines.

Grades vs. trendline direction. To investigate the second question (i.e., Was graph trendline direction related to class grade?), the number of ascending trendlines, descending trendlines, and horizontal trendlines were counted for each grade category (i.e., As, Bs, Cs, Ds, Fs). Table 2 displays the results.

These results indicate that of the 33 participants who earned As, 23 (or 70%) had graphs with ascending trendlines, and 7 (21%) had descending trendlines. Three of 33 participants (9%) had trendlines that were horizontal. Of those who earned Bs, 24 (or 75%) of the participants had graphs with ascending trendlines and 8 (or 25%) had descending trendlines. No participants who earned Bs had a horizontal trendline. Of those participants who earned a C in the class, two (67%) had ascending trendlines, and one (33%) had a descending trendline. No participant who earned a C had a horizontal trendline. There were no Ds earned in this class. The one student who failed the course had an ascending trendline.

Table 2
Direction of Trendlines Compared to Grade Category for 69 Students

Grade	Ascending	Descending	Horizontal	Total
A	23	7	3	33
B	24	8	0	32
C	2	1	0	3
D	0	0	0	0
F	1	0	0	1

Mode of trendline direction over the duration of the class. The third question—which, if any, graph trendline mode (ascending, descending, or horizontal) was more prevalent—was determined by comparing the number of student graphs that had ascending, descending, or horizontal trendlines. Of the 69 student graphs completed over the duration of the course, 50 displayed ascending trendlines, 16 displayed descending trendlines, and three displayed horizontal trendlines. Table 3 displays these results.

Table 3
Graph trendline mode for 69 students

Ascending	Descending	Horizontal	Total
50	16	3	69

Discussion

Classroom evaluation that is valid, reliable, and easy to conduct is critical to determining whether students are learning what has been taught (Summers, Beretvas, Svinicki & Gorin, 2005). This important concept applies equally to students in higher education as it does to students in other educational environments.

This preliminary investigation was conducted to gather data on whether participants' Accuplacer reading skills test scores predicted trendlines for CBM probe scores, whether CBM probe trendlines predicted participants' class grades, and whether ascending, descending, or horizontal trendlines were more likely to be demonstrated in student graphs.

In this investigation, there appeared to be little, if any, relationship between reading skills test scores and the direction of the probes' trendlines. Probe trendlines appeared to be relatively evenly distributed across reading skill test results, so there did not appear to be any relationship between Accuplacer reading skills test scores and students' graph trendlines. There appeared to be little correlation between Accuplacer reading scores and the slope of students' graph trendlines.

Data on the second question appears to be a bit more promising, but still did not, in the investigators' opinions, rise to a level of significance. Those participants earning passing (A, B, C) grades in the course had graphs with ascending trendlines (n = 49) more often than participants who got both passing grades and descending trendlines (n = 16). Lack of a more normal distribution of grades (most students received A's or B's, no student received a D, and only one student failed the course) discouraged speculation on possible relationships between trendline direction and grades. In addition, the graph of the one student who failed the class had an ascending trendline. The only relationship that appeared to be of any note was that students who passed the course were more likely to have graphs with ascending trendlines than graphs with descending trendlines.

Data on the third question relates to observations by Espin and Deno (1994-1995) as well as Espin and Foegein (1996) who found that a vocabulary matching task was a measure of learning on reading and content-area learning tasks. That most students' graphs (50 out of 69, or 72%) had ascending trendlines may indicate that individuals in the class improved their performance as the class progressed. Such a finding, if supported by subsequent research, might provide the basis for determining student progress on a formative basis. Determining formative progress might then afford the instructor opportunities to intervene with strategy training, increased support, and other forms of intervention for students not likely to pass the course without such assistance.

Limitations. Several limitations exist in this study. Of primary importance is the lack of a normal distribution of grades across the classes. Participants who earned an A, B, or C in the class had far more ascending graph trendlines than descending or horizontal ones, but the lack of Ds and paucity of Fs leaves open the question regarding what direction the trendlines might take.

Future research. Until recently, CBM procedures were not available for secondary school students. A modicum of data for using CBM in vocational training classes does exist (see Sticht, 1999), but none, apparently, directly applies to university students. Formative evaluation assessing university students' knowledge by graphing the outcomes of weekly CBM probes might help identify those students who are not learning the content-area knowledge at a high enough rate to ensure a passing grade. Instructors might propose additional support, such as small study groups or individual tutoring, to those students showing downward trends in their CBM probes. Using CBM at the university level might also help to reduce the dropout rate among freshman students by providing a more realistic picture of where such students stand in relation to a passing grade, as the class progresses (see Espin & Tindal, 1998, pp. 234 ff. for a discussion of such interventions), especially if flexible, intensive intervention were offered to potentially failing students.

More research appears to be needed to extend to the university level these assessment tools that have demonstrated usefulness at the elementary and secondary grade levels. Instructors—and students—have as great a need for indications of formative success or failure in a university class as do teachers and students in elementary, middle school, and high school settings.

References

Allinder, R. M. (1995). An examination of the relationship between teacher efficacy and curriculum-based measurement. *Remedial and Special Education*, *45*(2), 15-24.

Baker, S. K., & Good, R. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second grade students. *School Psychology Review*, *24,* 561-578.

Bean, R. M., & Lane, S. (1990). Implementing curriculum-based measures of reading in an adult literacy program. *Remedial and Special Education, 11*(5), 39-46.

College Entrance Examination Board. (2002). *Accuplacer.* New York: Author.

Council for Exceptional Children. (2005). Expert connection. In *TeachingLD: Information & resources for teaching students with learning disabilities.* Retrieved May 30, 2005, from *http://www.teachingld.org/expert_connection/cbm.html*

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37,* 187-192.

Deno, S. L., Marston, D., Mirkin, P. K., Lowry, L., Sindelar, P., & Jenkins, J. (1982). The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). Minneapolis:

University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED227129)

Espin, C. A., & Deno, S. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education, 14*, 47-59.

Espin, C.A., & Deno, S. (1994-1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20,* 121-142.

Espin, C. A., & Foegen, A. (1996). Curriculum-based measures at the secondary level: Validity of three general outcome measures for predicting performance on content-area tasks. *Exceptional Children, 62,* 497-514.

Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214-253). New York: Guilford Press.

Frank, A., & Gerkin, K. (1990). Case studies in curriculum-based measurement. *Education and Training in Mental Retardation, 25*, 113-119.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, *21*, 449-460.

Fuchs, L. S., & Fuchs, D. (1986). Curriculum-based assessment of progress toward long-term and short-term goals. *Journal of Special Education*, *20*, 69-82.

Fuchs, L. S., & Fuchs, D. (1988). Curriculum-based measurement: A methodology for evaluating and improving student programs. *Diagnostique, 14*, 3-13.

Fuchs, L.S., Fuchs, D. & Hamlett, C.L. (1994). Strengthening the connection between assessment and instructional planning with expert systems. *Exceptional Children, 61,* 138-146.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*, 20-29.

Gable, R. A., Arllen, N. L., Evans, W. H., & Whinnery, K. M. (1997). Strategies for evaluating collaborative mainstream instruction: "Let the data be our guide." *Preventing School Failure, 41*(4), 153-158.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.

Kim, D. (1998, April). *Specification of growth model and inter-individual differences for students with severe reading difficulties: A case of CBM.* Paper presented at the annual meeting of the Council for Exceptional Children, Minneapolis, MN. (ERIC Document Reproduction Service No. ED418553)

McDaniel, C. E., Evans-Hampton, T. N., Skinner, C. H., Henington, C., & Sims, S. (2002). An investigation of situational bias: Conspicuous and covert timing during curriculum-based measurement of mathematics across African American and Caucasian students. *School Psychology Review, 31*(4), 529-539.

Myers, D. G. (2004). *Psychology* (7th ed.). New York: Worth Publishers.

Naace. (n.d.). MAPE Random Number Generator [Computer software]. Retrieved February 15, 2005 from *http://www.humboldt.k12.ca.us/sobay_sd/district/randomiser.1/random.htm*

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single case research design and analysis* (pp. 15-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shinn, M. R. (Ed.) (1998). *Advanced applications of curriculum-based measurement.* New York: Guilford Press.

Shinn, M. R., & Baimonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: Guilford Press.

Shinn, M. R., Collins, V. L., & Gallagher, S. (1998). Curriculum-based measurement and its use in a problem-solving model with students from minority backgrounds. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 143-174). New York: Guilford Press.

Sticht, T. G. (1999). *Testing accountability in adult literacy education.* El Cajon, CA: Applied Behavioral & Cognitive Sciences.

Summers, J. J., Beretvas S. N., Svinicki M. D., & Gorin, J. S. (2005). Evaluating collaborative learning and community. *The Journal of Experimental Education, 73*(3), 165-188.

W. Corry Larson*, Ph.D., Assistant Professor of Special Education at the University of Maryland Eastern Shore, teaches both undergraduate and graduate education Special Education courses. He has conducted research in Curriculum-Based Measurement, Learning Strategies, and Direct Instruction, and is interested in a variety of topics related to these subjects. He can be contacted at* wclarson@umes.edu.