

## SMALL SCALE MEASUREMENT IN SOCIAL STUDIES

Luke Duesbery

San Diego State University  
Todd Twyman, Pacific University  
Kimy Liu, California State University, Stanislaus

DOI: <http://dx.doi.org/10.37500/IJESSR.2022.5418>

### ABSTRACT

This small-scale study investigated the use of a Concept Maze as an extension of Curriculum-based Measurement (CBM) to monitor the progress of student learning in a social studies classroom. Fifty 6th grade students participated in a three week conceptually based unit on Early Societies. All students were assessed with vocabulary CBMs and Concept Mazes on three separate occasions, at the end of each week. Results indicated that the vocabulary measures were neither related to the Concept Maze scores nor predictive of achievement on a criterion measure. Concept Mazes, however, were predictive of performance on the end of unit test. Mean growth patterns indicated that Concept Mazes were sensitive enough to monitor the learning progress of students with and without learning disabilities.

**KEYWORDS:** progress monitoring, curriculum-based measurement, CBM, social studies, concept maze, action research

We hope classroom instruction leads to student learning but, knowing instruction leads to learning is a different matter. To know if students are learning, many districts turn to results from federally mandated large-scale testing. However, the frequency of these tests, typically once each year, prevents them from being used to inform instruction. Further, the broad content sampling of these large-scale tests prevents them from being used to determine growth in more narrowly defined curricula. These problems are exacerbated in the content areas by less frequent administration, or in social studies where students are rarely, if ever, assessed. In response to these issues, a family of smaller progress monitoring-oriented curriculum-based measures (CBM) has evolved over the last 40 years. These CBMs are established indicators of student growth in reading, math, and writing, in the elementary grades, and have their greatest foothold in special education. Extending CBM into general education and the content areas has only recently begun to take shape. The purpose of this study was to investigate the predictive validity and sensitivity of a Concept Maze in a social studies curriculum for students with and without learning disabilities.

### Current Assessment Systems in Social Studies

Typical assessment systems teachers use, e.g. fact-based end-of-chapter tests, assume that students have mastered the content when a student demonstrates success (Frisbie, Miranda, & Baker, 1993).

Teachers also assume that tests are aligned with the textbook (Airasian, 1991), and are of high enough technical quality as to indicate content comprehension (Dorr-Bremme & Herman, 1986), though this is often not the case. End-of-chapter tests may fail to match the content of the objectives for its corresponding unit, comprise mainly multiple-choice and matching vocabulary or factual items (Ketterlin-Geller, McCoy, Twyman, & Tindal, 2006), entail minimal open-ended questions requiring fact-based responses (Frisbie, Miranda, & Baker, 1993), and provide little supporting evidence of their technical quality (Ketterlin-Geller, McCoy, Twyman, Tindal, 2003). Compounding the issue of perception is the fact that research has repeatedly shown that students with Specific Learning Disabilities (SLD) and low-achieving students do not succeed in factually based assessments (Prater, 1993; Schulte & Villwock, 2004), thus raising concerns about the validity of such assessments.

Further, the most limiting factor for teachers in recognizing student mastery is the timing of the tests themselves. Typically, tests are given at the end of the chapter without considering the progress of student learning. Grades are assigned, and the next chapter begins. In other words, performance on one test is independent of performance on another, and for all students, and particularly for students with SLD, summative types of classroom-based assessments may contribute to continued poor performance (Prater, 1993). Given this situation, there arises a need to introduce a systematic method of formatively assessing individual progress to counter mastery driven instructional models; Small scale tests, like CBM, can form the basis of such a system.

### **Curriculum Based Measurement**

In more than 40 years of empirical research, CBM has emerged as the leading assessment system to monitor learning progress. As originally conceived, CBM was designed to be a set of measures acting as critical indicators of general proficiency, allowing both special and general education teachers to administer interventions before students fail (Deno, 1985; Espin, Busch, Shin, & Kruschwitz, 2001). There are four basic tenets of CBM: (1) tests can be quickly and easily administered and scored, (2) are reliable from one administration to the next, (3) are valid indicators of the progression of student growth over time, and (4) results of individual progress can be easily charted and displayed (Deno, 1985).

Research demonstrates that CBM produces accurate, meaningful information about students' academic levels and are sensitive to student growth and improvement (Fuchs, 2004). The CBM score represents an individual's global level of competence in the domain, which can then be used by educators to "identify discrepancies in performance levels between individuals and peer groups, which helps inform decisions" about instructional interventions and accommodations (Deno, Fuchs, Fuchs, Marston, & Shin, 2001, p1). Although mastery measurement can demonstrate individual skill proficiency, it attributes performance to learning, which may or may not be the case. It is well documented in the literature that other variables contribute to learning, such as attendance, behavior, and homework completion. Unlike mastery measurement, CBM measures individual progress towards the curriculum domain as measured by a criterion.

Since its inception, CBMs have been used as progress monitoring tools to track student learning trajectories (Fuchs & Fuchs, 1993). As a progress monitoring tools, student learning is tracked through regular administrations of CBMs, providing the teacher with sufficient information to modify curriculum and instruction as needed. For example, if a student fails to meet a certain instructional goal, or is not progressing at the expected rate on three separate administrations, the teacher can refer to the types of errors that have been graphically displayed to modify the curriculum or instruction (Hintze, Christ, & Keller, 2002). As indicators of performance, correlations between CBM and criterion measures have ranged from .60 to .90, and test-retest and alternate form correlations are typically above .80 (Marston, 1989). Further, regular and ongoing use of CBM has provided classroom teachers with a reliable and valid system of monitoring student progress to improve their instructional planning and increase student achievement (Deno 2003; Foegen, Espin, Allinder, & Markell, 2001; Tindal, 1998). There are several large families of CBM including Easy CBM (Riverside), DIBELS (University of Oregon), and AIMSwebPLUS (Pearson).

### **The Typical Maze CBM**

As a formative measure, the maze task is a CBM that has been validated in reading at the elementary level (see Fuchs & Fuchs, 2002, for a review). Recently some have suggested that progress monitoring in social studies can be achieved with a simple vocabulary-matching techniques (Lembke et al, 2017, Byers et al, 2013). A typical Maze task is different from this approach. A Maze is a multiple-choice test where 10-16 words are deleted from an intact passage in a fixed ratio (every nth word), excluding articles and prepositions, and replaced with syntactically correct and semantically possible distracters. Students select the missing word from a list of words that best completes a sentence. Distracters are determined by level of comprehension and difficulty, and categorized by the following descending degree of difficulty from the correct option: (a) a subtle matter of degree, (b) a semantic understanding at the sentence level, where the distracters use the same part of speech as the omitted word, and (c) a syntactic understanding at the sentence level (nonsense words), where distracters were a different part of speech and are not meaningful to the sentence. The first and last sentences remain intact, and students are asked to circle the correct option. Items are generally scored dichotomously (Parker, Hasbrouck, & Tindal, 1992).

### **The Concept Maze CBM**

The Concept Maze borrows format and function of the typical maze, described above, and applies it to a more complex measurement paradigm by assessing content comprehension that focuses on the relationship between facts and concepts as the formalization of domain-specific conceptual knowledge (Alexander, Schallert, & Hare, 1991). We consider concepts to be part of a taxonomy of increasingly complex knowledge forms that consist of facts, concepts, and principles (Tindal & Nolet, 1995). In this taxonomy, facts, or one-to-one relationships between names, objects, places, or events, are the simplest forms of knowledge. Concepts are defined as one-to-two word abstractions that share a common set of defining characteristics or attributes (Klausmeier, 1990), to which factual

examples can be applied. Facts are critical to build in-depth conceptual knowledge in that they provide the example set of any given concept.

For example, revolution is a social studies concept. The defining attributes could be: (a) an oppressed majority; (b) powerful minority; and (c) a catalyst for rapid change. This concept and its conditions that must be present for any revolution to occur can be applied to any revolution, from the Russian to the sexual and technological revolutions. The explicitness of this instructional model has shown to have educational benefits for all students (Hollenbeck & Tindal, 1996; McCleery & Tindal, 1999; Twyman, 2003; Twyman, McCleery, & Tindal, 2004a; Twyman, McCleery, & Tindal, 2004b; Twyman & Tindal, 2005).

### **Emerging Research on the Concept Maze**

Three preliminary studies have investigated the effectiveness of the Concept Maze. In a measurement only study, Twyman & Tindal (2005) sought to differentiate the Concept Maze from the traditional reading comprehension maze. 240 students were randomly assigned to take a typical or Concept Maze. Results indicated that the Concept Maze was more difficult and more stable relative to the traditional maze. Performance on the Concept Mazes averaged .58 of a standard deviation below that on the traditional maze, and in terms of stability, alternate form reliability ranged from  $r = .19$  to  $.27$ , and test re-test correlations ranged from  $r = .44$  to  $.47$ . In a follow-up study, two biography Concept Mazes using different passages were randomly administered to 102 6th grade students. Alternate form correlations ranged from  $r = .54$  to  $.62$ , and test re-test correlations ranged from  $r = .51$  to  $.53$ . These results point to moderate reliability across concepts and populations.

Based on the above findings, Ketterlin-Geller, McCoy, Twyman, & Tindal (in press) implemented a series of Concept Mazes as formative instructional supports and developed data displays to facilitate teacher decision-making regarding individual student subject-matter knowledge. Eighty-four 6th grade students from four language arts/social studies block classes participated in a concept based unit on Life in Ancient Greece; the concept of civilization supported by attributes geography, religion, economics, government, and writing. Six different Concept Mazes were administered over the course of a 4-week instructional period. In this study, the internal consistency of the Concept Mazes were moderate, ranging from  $\alpha = .68$  to  $.77$ , and test-retest correlations ranged from  $r = .70$  to  $.95$ .

This study continues this line of research with two research questions. Espin and colleagues (Espin & Deno, 1993a, 1993b, 1994-1995; Espin & Foegen, 1996; Espin, Shin, & Busch, 2005) have concluded that vocabulary CBMs are reliable and valid indicators of general academic performance at the secondary level. We feel, however, these CBMs miss a critical component in content curriculum. Thus, similar to Twyman and Tindal (2005), we sought to further explore the distinguishing features of the Concept Maze compared to vocabulary CBMs. Second, considering that this study employed an emerging instructional paradigm focusing on explicitly organizing the

curriculum around concepts and attributes, we sought to investigate the predictive validity of the Concept Maze for monitoring progress.

## **METHOD**

### **Participants**

Two groups of 6th grade students ( $n = 50$ ) from one rural middle school participated in the study. Participants were primarily Caucasian (80%), of low socioeconomic status (90% eligible for free/reduced lunch), and similar with respect to gender (23 male and 27 female). Of the 50 students, 16 received special education services in reading and/or writing. All students were included in the study, and there was no attrition.

### **Concept-Based Instruction**

Based on an 8th grade state standard, the topic of Early Societies was chosen by the teacher. In conjunction with the second author, the teacher determined that the Early Societies covered in the text had three critical attributes. The attributes were that Early Societies (a) joined together to meet their basic needs, (b) had a means of communicating with each other, and (c) created artwork as a mode of expression.

The textbook was organized into two separate chapters highlighting Hunters and Gatherers, and Early Farmers. Treating these as separate chapters, however, does not promote an understanding of the major characteristics of Early Societies, because the links between the attributes remain embedded in the text and become instructionally dependent. The concept/attribute structure helps bridge the facts to the concept across chapters. Placing individual facts into a common set of attributes of Early Societies helps the learner distinguish the important commonalities and differences among the Hunters and Gatherers and Early Farmers to help them strategically use that information when explaining their influence on civilization. Students can then apply those same attributes to a new context in evaluating any other Early Society of other times or eras.

### **Measures**

This study used two formative measures, a Concept Maze and vocabulary matching CBM, and one criterion measure, a graphic organizer. Each formative measure was administered as a package, and the criterion measure was administered as a posttest.

**Concept Mazes.** All students were administered three Concept Mazes (once per week). The Concept Maze is similar in format to traditional the traditional Maze, however, their construction protocols differ. The Concept Mazes were taken from three randomly sampled passages found in a different textbook and created according to the following procedures: (a) each passage was reviewed by the teacher and second author and considered to be representative of the topic under study, (b) a full range of possible attribute examples were identified, (c) the density of each attribute example was

equal and inserted at approximately every 25th word interval, (d) reading difficulty and length were adjusted to a 6th grade reading level, (e) each attribute example was bolded and attribute choices were placed in brackets beside the bolded item, (f) the first and last sentences were left intact, and (g) objective scoring rules were applied. See Figure 1 for a sample item.

### Figure 1. Example Concept Maze Item

Early people were mainly hunters who knew nothing about raising animals or farming. They **lived in bands of about 20 people** (needs/language/art). They **ate whatever animals they killed or caught** (needs/language/art), and whatever fruits and plants they found growing. Meat was their main source of food.

As the organization of each passage differed, the numbers of items differed from passage to passage. Concept mazes 1 and 2 each consisted of 14 items with attribute examples evenly distributed. Maze 3 consisted of 19 items with attribute examples evenly distributed.

Vocabulary CBMs. Following the current vocabulary CBM development and administration protocol suggested by Espin, et al. (2001), the teacher and the second author identified 48 important vocabulary words related to the concept and attributes, including the concepts and attributes themselves. From this pool, three forms were created. Each vocabulary measure consisted of 16 vocabulary words randomly taken from the pool of 48 words. In each form, the vocabulary terms were placed vertically in random order on the left side of the page, and matching definitions were placed vertically in random order on the right.

Criterion measure. In keeping with the conceptual nature of the instruction, a graphic organizer was used as the criterion measure to assess students' level of conceptual understanding. This type of measure has emerged as a viable alternative to textbook-based tests (Jonassen, 2000; Koul, Clariana, & Salehi, 2005; Ruiz-Primo & Shavelson, 2000). Students were given a blank graphic organizer, identical to the one they used as their note-taking device, and were asked to record the concept label, attributes, and as many examples per attributes as possible. Scores were calculated by the number of answers correct when compared to a list of all possible answers derived from the chapter (68 total).

The concept label was worth three points, two points per attribute, and one point per factual example, corresponding to the relative importance of each component. This scoring method is consistent with Rye and Rubba's (2002) scheme for relationship weighted scoring. After coding and shuffling the papers, two researcher-trained graduate students scored the graphic organizer. Pearson correlation coefficients indicated that inter-rater reliability was high,  $r = 0.90$ .



**Procedure**

The study lasted approximately 3 weeks, during which time the participating teacher covered a two-chapter unit on Early Societies from the district's adopted textbook *The World* (Boehm, et al., 2000). All instruction was conducted during the same 45-minute periods daily.

As a conceptual approach to learning was new to the students, the first two days comprised an introduction to the purpose and sequence of instruction and defining the concept of Early Societies and its attributes. To facilitate a discussion of the concept and attributes, the teacher provided students with definitions of each, and then asked each class to read a section of the text and brainstorm examples that would fit into each attribute. The teacher facilitated discussion by categorizing each example into concept/attribute columns on the board. Each student transcribed the examples into a graphic organizer designed to be used as their notes. The teacher used several different reading strategies, e.g., K-W-L (Ogle, 1986), Think-Pair-Share, as part of her activities, and for the remainder of the study, she began each lesson with a review of the examples learned from the previous day, and ended the lesson with a review closure activity.

**Fidelity of Implementation**

During the training phase, the teacher attended two workshops designed to familiarize her with the method of instruction, at which time the lead author and teacher developed the unit designed for this study. As the intervention proceeded, the second author and a trained graduate student observed both classes twice per week to check for the explicit usage of concepts and attributes, which were recorded onto a checklist. Using a momentary time sample, one of the two observers recorded teacher use of the concept or attribute labels, and examples of the attributes every 15 seconds over a twenty minute interval. Two inter-observer checks were conducted. The reliability was 1.0. Totals from the observations indicated that the teacher referred to one of the critical concept elements approximately every 1.5 minutes. The totals were similar across classes and lessons.

**RESULTS**

The first research question attempted to distinguish the Concept Maze from vocabulary measures. As Table 1 indicates, inter-correlations between vocabulary measures were moderate and ranged from  $r = .46$  to  $r = .67$ ,  $p < .05$ . Inter-correlations between Concept Mazes are also moderate, ranging from  $r = .41$  to  $.51$ ,  $p < .05$ . Vocabulary CBMs were not significantly related to any of the Concept Mazes. Regression analyses indicated that the vocabulary measures did not appear to predict performance on the criterion measure. In contrast, performance on the Concept Mazes significantly predicted between 18-25% of the variance in the end of the unit graphic organizer test ( $p < .05$ )

**Table 1. Inter-correlations between the Concept Maze and Vocabulary Measures**

|              | Maze 1 | Maze 2 | Maze 3 | Vocabulary 1 | Vocabulary 2 | Vocabulary 3 |
|--------------|--------|--------|--------|--------------|--------------|--------------|
| Maze 1       |        | .41*   | .51*   | .21          | .22          | .25          |
| Maze 2       |        |        | .44*   | .13          | .18          | .28          |
| Maze 3       |        |        |        | -.07         | .08          | -.07         |
| Vocabulary 1 |        |        |        |              | .66**        | .46**        |
| Vocabulary 2 |        |        |        |              |              | .67**        |
| Vocabulary 3 |        |        |        |              |              |              |

\*  $p < .05$ , \*\*  $p < .01$

The second research question addressed the sensitivity of the Concept Mazes to monitor learning progress. As Table 2 indicates, growth was evident across all three Concept Maze administrations, with the most marked growth occurring by the end of the second week of the three week instruction period.

**Table 2. Means and Standard Deviation (%)**

|      | Maze 1 | Maze 2 | Maze 3 |
|------|--------|--------|--------|
| Mean | 53     | 61     | 68     |
| SD   | 14     | 9      | 7      |

A repeated measures analysis of variance was conducted to evaluate the effect of educational classification on conceptual knowledge represented by performance on the Concept Maze. Results showed no statistical interaction effect between students with and without learning disabilities, indicating that students with LD performed comparable to their general education peers.

## DISCUSSION

In social studies, content can be seen to revolve around common concepts, and concepts can serve sources for stable and valid measurement in an otherwise poor assessment system. Though there is ample evidence supporting CBM in reading and math at the elementary levels (Deno, 2003), subject areas with more diverse content, such as social studies and science, have little research supporting such responsive measures. Thus, the intent with this research was to present the Concept Maze as a viable CBM of content area curriculum.

The first research question attempted to distinguish Concept Maze from vocabulary CBMs. Results from this study support this supposition that they are indeed distinct from one another. Student performance on the vocabulary CBM was not correlated with their performance on the Concept Maze. This suggests the two CBMs were measuring something different. Furthermore, examining the



predictive power of the vocabulary CBM and Concept Maze measures on the criterion measure revealed that they were, again, different. The vocabulary measure explained very little variance on the criterion measure, whereas the Concept Maze consistently explained a significant amount of the variance on the same criterion. In light of the literature regarding the murky interplay between vocabulary acquisition and content learning (Blachowicz & Fisher, 2000; Nagy & Scott, 2000), these results could provide the field with a value-added assessment tool. That is to say, given that vocabulary CBMs measure something distinct from Concept Mazes, administering both might be complementary. Clearly more research is warranted to address this notion.

In our study students with and without disabilities grew in terms of performance on the Concept Maze. The growth was expected given the nature of instruction delivered, and serves to demonstrate the Concept Maze can be sensitive to the learning growth of all students.

## CONCLUSIONS

In a progress monitoring model, we increase the frequency of measurement and target key achievement skills with reliable CBMs. By combining CBMs with instructional intervention, learning deficits can be identified based on non-responsiveness. This binding of instruction and progress monitoring has most recently been called Response to Intervention (RTI) (Gresham, 2002). In special education, progress monitoring instruments are routinely used to diagnose deficiencies, but more recently have been called on as a tool for identifying learning disabilities (IDEA, 2004). The results from this study suggest that Concept Mazes might be added to the family of CBMs, including vocabulary CBMS, and might be used for all students, regardless of learning disability.

We know that multiple-measure models, as in a progress-monitoring system, increase the precision of estimating a student's learning status and growth rate; data are based on the entire collection of measures, not a single datum (Gresham, 2002). From a classroom perspective, this suggests that just using end-of-chapter tests or comparable summative measure might not necessarily lead to valid inferences about student learning.

On a more general level, results from this study provide further support for using CBM as measures of change in content areas. As argued by Fletcher, et al. (2005), CBM could be used in combination with other measures in the domain to determine to what extent students are on par or discrepant from their peers at a single point in time, as well as to what extent students are progressing both individually and relative to their peers. Students who are discrepant both in performance and progress would be those most in need of intensive intervention, while students who show adequate growth could be considered to be removed from academic support in this domain.

**REFERENCES**

- Airasian, P. W. (1991). *Classroom assessment*. New York: McGraw-Hill.
- Alexander, P., Schallert, D., & Hare, V. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61(3), 315-343.
- Blachowicz, C. L. Z., & Fisher, P. (2000). Vocabulary instruction. In M L, Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 503-524), Mahwah, NJ: Erlbaum.
- Beyers S., Lembke E. & Curs B. (2013). Social studies progress monitoring and intervention for middle school students, *Assessment for Effective Education*, 38(4).
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Deno, S. L. (2003). Curriculum-based measures: development and perspectives. *Assessment for Effective Intervention* 28(3 & 4): 3-12.
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using Curriculum-Based Measurements to establish growth standards for students with learning disabilities. *School Psychology Review*, 30(4).
- Dorr-Bremme, D. W., & Herman, J. L. (1986). *Assessing student achievement: A profile of classroom practices*. Los Angeles: Center for the Study of Evaluation, University of California.
- Espin, C. A., & Deno, S. L. (1993a). Content-specific and general reading disabilities of secondary-level students: Identification and educational relevance. *The Journal of Special Education*, 27, 321-337.
- Espin, C. A., & Deno, S. L. (1993b). Performance in reading from content-area text as an indicator of achievement. *Remedial and Special Education*, 14(6), 47-59.
- Espin, C. A., & Deno, S. L. (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). Validity of three general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children*. 62, 497-514.
- Espin, C. A., Busch, T., Shin, J., & Kruscwits, R. (2001). Curriculum-based measures in the content areas: Validity of vocabulary-matching measures as indicators of performance in social studies. *Learning Disabilities Research & Practice*, 16, 142-151.

- Espin, C. A., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in the content areas: Vocabulary matching as an indicator of progress in social studies learning. *Journal of Learning Disabilities, 38*(4), 353-363.
- Fletcher, J. M., Francis, D. J., Morris, R. D., & Lyon, G. R. (2005). Evidence-based assessment of learning disabilities in children and Adolescents. *Journal of Clinical Child and Adolescent Psychology, 43*(3), 506-522.
- Foegen, A., Espin, C.A., Allinder, R.M., & Markell, M.A. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *Journal of Special Education, 34*, 226-236.
- Frisbie, D. A., Miranda, D. U., & Baker, K. K. (1993). An evaluation of elementary textbook tests as classroom assessment tools. *Applied Measurement in Education, 6*(1), 21-36.
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based research. *School Psychology Review, 33*(2), Retrieved July 7, 2005, from Academic Search Premier database.
- Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment nonresponders. *Peabody Journal of Education, 77*(2), 64-84.
- Fuchs, L.S., & Fuchs, D. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1).
- Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 467-564). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hintze, J.M., Christ, T.J., & Keller, L.A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*, 514-528.
- Hollenbeck, K. & Tindal, G. (1996). Teaching law concepts within mainstreamed middle school social studies settings. *Diagnostique, 21*(4), 37-58.
- Jonassen, D. (2000). *Computers as mindtools for schools*. Englewood Cliffs, NJ: Prentice-Hall.
- Ketterlin-Geller, L.R., McCoy, J.D., Twyman, T., & Tindal, G. (2003). How do critical thinking measures fit within standards-based reform? *Assessment for Effective Intervention, 28* (3&4), 37-48.

- Ketterlin-Geller, L.R., McCoy, J.D., Twyman, T., & Tindal, G. (2006). Using a Concept Maze to Assess Student Understanding of Secondary Content. *Assessment for Effective Intervention*, 31(2), 39-50.
- Klausmeier, H. J. (1990). Conceptualizing. In B.F. Jones & L. Idol (eds.), *Dimensions of thinking and cognitive instruction* (pp. 93-138). Hillsdale, NJ: Erlbaum.
- Koul, R, Clariana, R. B., & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. *Educational Computing Research*, 32(3), 227-240.
- Lembke, E., Allen, A., Cohen, D., Hubboch, C., Landon, D., Bess, J., & Bruns, H. (2017). Progress Monitoring in Social Studies Using Vocabulary Matching Curriculum-Based Measurement, *Learning Disabilities Research and Practice*, 32(2).
- Marston, D. (1989). Technical adequacy of reading measures. In M. Shinn (Ed.), *Curriculum-based measurement* (pp. 120-154). New York: Guilford.
- McCleery, J. A., & Tindal, G. (1999). Teaching the scientific method to at-risk students and students with learning disabilities through concept anchoring and explicit instruction. *Remedial and Special Education*, 20(1), 7-18.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M L, Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III), Mahwah, NJ: Erlbaum.
- Ogle, D. M. (1986). K-W-L: A teaching model that develops active learning of expository text. *The Reading Teacher*, 39, 564-570.
- Parker, R., Hasbrouck, J., & Tindal, G. (1992). The maze as a classroom-based reading measure: construction methods, reliability, and validity. *The Journal of Special Education*, 26(2), 195-218.
- Prater, M. A. (1993). Teaching concepts: Procedures for the design and delivery of instruction. *Remedial and Special Education*, 14(5), 51-62.
- Ruiz-Primo, M., & Shavelson, R. (2000). Psychometrics of assessing student understanding. In J. Mintzes, J. Wandersee, & J. Novak, (Eds.) *Assessing science understanding* (pp. 303-341). San Diego, CA: Academic Press.
- Rye, J. A., & Rubba, P. A. (2002). Scoring concept maps: An expert map-based scheme weighted for relationships. *School Science and Mathematics*, 102(1), 33-44.
- Schulte, A. C., & Villwock, D. N. (2004). Using high-stakes tests to derive school-level measures of special education efficacy. *Exceptionality*, 12(2), 107-126.



- Tindal, G. (1998). Assessment in learning disabilities. In J. Torgesen & B. Wong (Eds.). *Perspectives in Learning Disabilities*. San Diego, CA: Academic Press.
- Tindal, G., & Nolet, V. (1995). Curriculum-based measurement in middle and high schools: Critical thinking skills in content area. *Focus on Exceptional Children*, 27(7), 1-22.
- Twyman, T., & Tindal, G. (2005). *Extending curriculum-based measurement in the middle/secondary schools: The technical adequacy of the Concept Maze*. Unpublished manuscript.