Scientific
Research
Publishing

Curriculum Based Measurement Maze: A Review

Sotiria Tzivinikou, Athanasios Tsolis, Dimitra Kagkara, Sevasti Theodosiou

Department of Special Education, University of Thessaly, Volos, Greece
Email: sotitzivi@uth.gr, tsolis.ath@gmail.com, dimitrakagka-ra8@gmail.com, theodosiou@uth.gr

Abstract

The maze is a type of assessment tool for reading in the context of Curriculum-Based Measurement (CBM). Over the last three decades, the maze, with the contribution of technology, is evolving into an assessment tool that can be automatically administered and scored. The objective of this study was to review the literature research on the maze. Sixty-three studies and other sources have been reviewed, and it was found that it is a technically adequate measurement for universal screening and monitoring progress and has a high correlation with reading achievement tests. Additionally, the maze seems to be a time-efficient and sensitive tool, included in most CBM tests. However, some significant issues emerging from the review need further investigation. These issues are related to the reading maze measures, its instructional utility as a progress-monitoring tool, and the determination of the maze's distractors.

Keywords

Curriculum Based Measurement, Reading Assessment, Reading Achievement, Maze, Technology, Universal Screening, Progress Monitoring, Measuring Growth

1. Introduction

Reading assessment is a key procedure in every educational system, especially when the educators' objective is reading proficiency in children with learning disabilities. The first step in assessment is to identify students at risk of reading failure through screening and refer them for further diagnosis (Tzivinikou, 2018). Reading difficulties are a characteristic of many children with learning disabilities, due to numerous cognitive deficits that may affect their reading ability (Gersten, Fuchs, Williams, & Baker, 2001). Instruction or remediation for child-

ren with reading difficulties is a complex procedure that always begins with assessment (Busch & Reschly, 2007; Coyne & Harn, 2006; Coyne, Zipoli, & Ruby, 2006; Stecker, Lembke, & Foegen, 2008).

Assessment procedures that serve diagnosis and performance measurement purposes, should be based on acceptable cognitive models, and be causally related to teaching (Leighton & Gierl, 2011). Assessment results must be understood by students, teachers, and parents and be clearly communicated to stakeholders by explaining how these results are going to be used in guiding remediation or instruction (Leighton & Gierl, 2011).

In the last few years, with the technology up growth, several assessment methods have been proven useful for automated administration and scoring, thus eliminating time constraints for educators. For example, computer-adaptive testing (Linacre, 2000) or scaffolding and integrated assessment can be used in computer-assisted learning environments (Beale, 2005).

Almost forty years after its first appearance, Curriculum Based Measurement (CBM) seems to have evolved in an appropriate method for automated administration, scoring, and representation of students' progress (Fuchs, 2004). With several diagnostic, screening, and progress monitoring methods developed so far, the latter seems to be the most challenging to create (Fuchs, 2004).

In this context, we investigated an assessment method for reading that can be used as formative assessment to give educators vital signs of students' performance and provide them with the necessary data for the decision-making process. CBM maze seems to fulfill the above as a General Outcome Measurement (GOM) for reading and can serve educators as a universal screening and progress-monitoring tool.

Therefore, the objective of the present study was to conduct a literature review of an assessment tool for reading, the maze, which can be used for universal screening and progress monitoring. More specifically, the research questions of the study are listed below:

Q1: Why was the maze initially developed and what is the maze's history?
Q2: Is the maze technically adequate for measuring students' growth?
Q3: Can the maze be used for universal screening for reading?
Q4: Which components of reading does the maze measure?
Q5: How often could the maze be used for measuring growth in reading?
Q6: Do teachers and students prefer the maze as a measurement method?

2. Procedure

Regarding the research process, the following criteria and search strategies pursued in information sources.

For the retrieval of research articles, books, and technical manuals, three criteria were set: 1) to have been published from 1974 to 2020, 2) to concern Curriculum Based Measurement Maze and 3) to have been published in scientifically recognized journals and books. The search was performed in jounals.sagepub, sciencedirect tandfonline, onlinelibrary.wiley, jstor and google scholar databases

in order to locate the appropriate literature. The keywords that were used to refine the research were: Curriculum Based Measurement, reading assessment, reading achievement, maze, technology, universal screening, progress monitoring, measuring growth. The initial search resulted in a collection of 100 sources including research articles, books, and technical manuals. Initial findings were studied in terms of title, abstract, results and discussion, which narrowed our collection to 63 sources that matched our topic interests.

3. Curriculum Based Measurement

Deno (1985) highlighted at an early stage the need to create easy-to-use and reliable tools for measuring and monitoring students' progress, which were mainly based on the informal observations of their teachers. In this context, the CBM measurement can be used both in general and special education, aiming universal screening in order to identify students at risk and evaluate intervention programs (Deno & Fuchs, 1987; Hintze, 2009; Hintze, Wells, Marcotte, & Solomon, 2018).

CBM's main difference from traditional reading tests is that it draws the materials of the measurements from the student's curriculum (Deno & Fuchs, 1987). At least this was the case in the early stages of its development where teachers constructed the tests each time using texts from school textbooks (Wayman, Wallace, Wiley, Tichá, & Espin, 2007). This choice offered high face validity, although it lacked validity and reliability (Shinn & Shinn, 2002b). Research on CBM started in 1977 by the University of Minnesota and its primary purpose was to develop measurement and evaluation tools that could be administered at short intervals to improve the educational decision-making process (Deno & Fuchs, 1987).

A CBM tool for reading should be valid and reliable, simple, and effective, produce easy, understandable, and interpretable results, and be cost-effective (Deno, 1985). In a complex skill such as reading, the measurements of a CBM tool should provide information on the progress of reading comprehension (Deno, 1985).

With the spread of response to intervention model (RtI) it became widely known, because it was considered an ideal method for detecting students at risk, in order to monitor progress and make data-based decisions on measuring the response to intervention (Hintze, 2009).

A considerable amount of literature has been published about CBM and its contribution to monitoring student progress is highlighted as their performance is assessed by multiple measurements over short periods of time and the assessment is designed to detect small changes in performance (Chung, Espin, & Stevenson, 2018).

CBM is necessary to fulfill specific traditional psychometric characteristics, such as the various types of validity and reliability to be considered technically adequate. It must meet some of the traditional psychometric characteristics, such

as the various types of validity and reliability. However, the traditional psychometric characteristics may not be sufficient to be considered an adequate assessment for measuring students' progress. It is necessary to explore additional technical features, such as students' growth across grades, Standard Error of Measurement (SEM), Standard Error of Estimate (SEE), as well as to ensure acceptance by teachers (Fuchs & Fuchs, 1992).

4. Curriculum Based Measurement in Reading

It is known that students with reading difficulties present slower development in reading skills. Thus, measurements must be made with tests that have high psychometric characteristics, mainly with high accuracy and objectivity to detect small changes in performance (Parker, Hasbrouck, & Tindal, 1992). In addition, these tools should be easy to administer, score, and interpret and have high face validity to gain popularity in the educational community, since teachers are the ones who use them regularly (Parker et al., 1992).

When CBM got popular and widely used as a method for universal screening and monitoring progress in 1985 (Deno), Oral Reading Fluency (ORF) played a key role in CBM tools as it was a commonly accepted measure for extracting a general outcome in reading ability (Fuchs & Fuchs, 1992); an approach still supported by Van Norman et al. (2018).

The universal consensus on the need for effective intervention strategies, highlighted the necessity to adopt practices to continuously monitor progress. In this context, for the last three decades, two types of measurements and tests exist: the Specific Subskills Mastery Measurement (SSMM), and the General Outcome Measurement (GOM) (Fuchs & Deno, 1991).

These tests are short and sensitive to changes in student performance and they are causally related to the educational decision-making process by providing data within the problem-solving model (Ball & Christ, 2012). They are considered to bridge the gap between traditional and more rigorous assessment methods in relation to teachers' informal observations, thus creating an innovative method for planning and making educational decisions, as well as monitoring progress (Fuchs & Deno, 1991).

In SSMMs, reading is "divided" into specific sub-skills, which are prioritized and for which individual tests are formed consisting of both the teaching goal and specific subskill measurement for monitoring progress (Hintze, Christ, & Methe, 2006). These measurements may emphasize a lot on a specific skill and not on the overall students' skills (Fuchs & Deno, 1991; McConnell & Wackerle-Hollman, 2015). As initially reported by Fuchs and Deno (1991) but also more recently by Ball and Christ (2012), these measurements are more suitable for short-term interventions, as they are not appropriate for generating the students' general overview and developmental level. However, it seems that a combination of the two approaches may be used by exploiting their positive aspects (Ball & Christ, 2012; Van Norman, Maki, Burns, McComas, & Helman, 2018).

On the other hand, GOMs seem to be more appropriate to formulate long-term teaching objectives, give teachers the freedom to try different teaching approaches and strategies, while providing a dependent variable for databased decision making over the intervention's effectiveness (Fuchs & Deno, 1991; Shinn & Shinn, 2002a; Van Norman et al., 2018).

ORF measure is considered as the most critical measurement in R-CBM since 1982 (Deno, Mirkin, & Chiang) and therefore the existing literature on ORF, over the last 10 years, focuses particularly on issues such as its utility for reading difficulties identification in other languages (Protopappas & Skalumbakas, 2008), probe equivalence (Christ & Ardoin, 2009), the way of graphically representing progress (Christ, Silberglitt, Yeo, & Cormier, 2010), whether R-CBM measurements can be used as indicators on the impact of teaching methods used by teachers (Petscher, Cummings, Biancarosa, & Fien, 2013), how to use the data resulting from monitoring the progress for the educational decision-making process (Burns et al., 2017; Van Norman & Christ, 2016) and the degree of correlation with reading comprehension (Shin & McMaster, 2019).

Although ORF provides useful information regarding reading development and is included in most CBM criteria, alternative types of reading measurements were sought, which could be administered in groups and in timely manner (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Reschly, Busch, Betts, Deno, & Long, 2009; Shinn, Knutson, Good, Tilly, & Collins, 1992).

The administration, scoring and representation of the data could be automated with technological assistance, the measurement results could have higher degree of correlation with reading comprehension and demonstrate higher face validity in order to be accepted by CBM users as a measure for reading comprehension (Fuchs & Fuchs, 1992). Furthermore, this test should meet the psychometric characteristics of validity and reliability and be sensitive enough to detect small changes in student performance.

5. Why Was the Maze Initially Developed?

The maze was developed in the early 1970s to improve reading assessment done with cloze procedures and was intended to assess students of low socioeconomic background, English Language Learners (ELLs) and students with reading difficulties (Parker et al., 1992). One of the first "official" bibliographic references was made by Guthrie, Seifert, Burnham, and Caplan (1974), who sought an alternative method of assessment, with the following characteristics: a) high validity and reliability, b) easily constructed by teachers, using materials similar to those of teaching c) short administration time and d) able to provide more data for reading comprehension compared to ORF. One of the main arguments they used was that there were students with particularly high performance in ORF and low reading comprehension performance; these students were known as "word-callers" (Guthrie et al., 1974). These students needed an assessment tool which would easily identify them and offer them the appropriate intervention to

improve reading comprehension. This issue first appeared in 1974 (Guthrie et al.) and is still of concern to the research community, since it lacks sufficient evidence, whether the maze is a better measuring tool than the ORF in identifying students with such difficulties (Shin & McMaster, 2019).

The maze may be used alternatively for selecting appropriate texts for reading comprehension instruction (Guthrie et al., 1974). It is reported that students with difficulty in comprehension made significant progress, as the level of difficulty of the texts increased on the appropriate timeframe (Guthrie et al., 1974). A grade point average of just over 65% is suggested for selecting the appropriate text to be taught and going to a more difficult level when the student reaches 85% - 100% (Guthrie et al., 1974).

6. Five Important Stages of Development in the Maze's History

The alternative proposed by Guthrie et al. (1974) is described as a maze procedure, in which any text from a book can be used. Every 5th or 10th word is replaced with three alternatives. One of them is correct, one belongs to the same grammatical category (i.e. verb), and one does not belong in the same category. Students read the text silently and choose the answer they think is correct. Then the students' correct answers are scored and converted into a percentage, and this is considered to be the student's level of understanding in the specific text. For example, in a text with 20 blanks, if a student has 15 correct answers, then it is considered that he/she can understand this text by 75%. The test was not timed, and the texts were about 120 words each, categorized by a readability formula. Students should not be familiar with the texts or the vocabulary used prior to the administration.

The maze was used as a tool to evaluate Basic Academic Skills Samples (BASS; Deno, Maruyama, Espin, & Cohen, 1989). There were texts of about 260 words in which every 7th word was deleted, and in its place, there were three possible options (distractors). Distractors were not relevant with the meaning of the text. To check the randomness of the students' answers a discontinue rule was used in the three incorrect answers and the students' score was derived from the correct answers in three texts.

While looking for tools to draw a general outcome in reading within the CBM that would meet the criteria for automated assignment and scoring, Fuchs and Fuchs (1992) found the maze as the most appropriate method for measuring progress. It was also appropriate because it was more acceptable to teachers who used CBM tools as a measure for reading comprehension, rather than ORF. They used texts of about 400 words to construct the tests, which included complete stories to facilitate meaningful reading and were categorized by readability formula to match grade's level.

Each maze is constructed as follows: The first and last sentences remain intact, and then every 7th word is deleted and replaced by one of three possible options

referred to as distractors. Distractors should have the same number of letters or differ by one. They should not 1) match the sentence semantically 2) be rhyming with the correct choice 3) look like the correct one orally or visually 4) be non-words 5) require students to read more than 1.5 line in order to exclude it as an option 6) not be so difficult in vocabulary that it does not have the possibility to be considered a non-word. The students have 2.5 minutes at their disposal and the administration is done using a computer. These rules for selecting distractors are followed in most maze-based surveys, except when researchers wish to increase the difficulty of the test (Shinn & Shinn, 2002b).

Aimsweb (Shinn & Shinn, 2002b) uses texts from 150 to 400 words. As in the development process described by Fuchs and Fuchs (1992), the first sentence of the text remains intact and then every 7th word is replaced by one of three possible options. The differentiation they introduce regarding distractors is a simplified selection process. One word is considered to be a "near" distractor, i.e. it belongs to the same grammatical category as the correct one, but does not match the meaning of the sentence while the other is a "far" distractor, i.e. it belongs to another grammatical category and is chosen randomly from the text without matching the meaning of the sentence. The score derives from the students correct minus incorrect answers.

In the latest version of DIBELS 8th (University of Oregon, 2020), the maze is designed to be administered from the 2nd to the 8th grade. The process of text development is the same as the ORF process. That is, texts were written by several experienced writers from different US states who had a different background. Emphasis was placed on making the texts interesting for the students and appropriate for each grade. The texts for the maze, however, are longer so that they can accurately measure the reading skills of children with better reading fluency and comprehension. The main difference with the maze, as described in previous stages, lies in the choice of distractors and the visual formatting of the test, utilizing research results that suggest that large series of texts can be difficult to read and understand.

The final score of the test is not only based on the correct answers of the students but is also based on the difference of the correct ones minus the incorrect answers divided by 2. Then this number is converted into an equated scale score with an average of 100 and a standard deviation of 10, to reduce the effect of the different difficulty level on each test. In this way a higher score from test to test will be due to the improvement of the students and not to differences in the text difficulty.

Regarding the construction of the test, the first and the last sentence of the text remain intact to allow the better consolidation of the meaning of the text. In the 2nd grade, the first two sentences and the last one remains intact. Starting with the 3rd word of the second sentence, every 7th word was deleted with some adjustments. If it was a proper noun or a number, then the 8th word was deleted. If the 7th word was a specialized term it was deleted, unless previously explained

in the text. Distractors could not start with the same letter as the correct choice, or have two letters different from the correct choice, although this rule became more relaxed for the higher grades (after the 5th grade). Moreover, the words should be grammatically correct if they were verbs, nouns, adjectives or contraction, excluding all other parts of speech, such as articles, where grammatical correctness was not necessary, as it seemed to lead to the same "distractors". Also, different types of the same word could not be selected. Finally, from 5th grade onwards one of the distractors should be semantically like the correct word, to fit the meaning of the sentence but not the text.

7. Is the Maze Technically Adequate for Measuring Students' Growth?

Technical features of slope are essential for CBM because they provide the students' actual progress in an academic domain and seem to be of the most challenging issues for CBM researchers and developers (Fuchs, 2004). To measure students' growth in reading comprehension, multiple measurements at short intervals are required in order to demonstrate the change of students' performance (Fuchs & Fuchs, 1992; Shin, Deno, & Espin, 2000).

In Guthrie et al.'s (1974) effort to investigate the maze's traditional psychometric characteristics, a high degree of correlation was found (0.82) with national standardized reading comprehension achievement test for 2nd grade students (Guthrie et al., 1974). They also compared the performance of students with and without learning disabilities, in terms of reading fluency and text maze. The results showed that students without difficulties have a similar performance (0.99) in ORF (89.54%) and in the text maze (89.23%), while students with learning difficulties had greater differences between the two tests. On average, 80.92% is in ORF and 60.62% in the text maze, with a correlation coefficient of 0.72 (Guthrie et al., 1974).

Jenkins and Jewell (1992) investigated concurrent validity across grade levels from 2nd to 6th grade of BASS reading test, which uses the maze as a measuring tool for reading, with nationally standardized reading achievement tests, in 413 children, and found a similar correlation of 0.85. Muijselaar, Kendeou, de Jong and van den Broek (2017), also found a high degree of correlation, from 0.69 to 0.79, between the maze and national achievement test, through 4th, 7th, and 9th grade.

Shin et al. (2000) examined the maze's validity, reliability, and ability to measure students' growth, in 43 2nd grade students with monthly measures during a school year. Their findings were equally encouraging, as a statistically significant relationship was found between the average growth of the monthly maze measurements and the state achievement tests. Alternate form reliability for 3-minute maze with intervals from 1 to 3 months, with an average rate of 0.81. Chung et al. (2018) found slightly lower scores, from 0.67 to 0.83 for the maze administrations lasting 2 minutes over a week and recorded an average weekly progress of

0.16 words for 7th grade students. Tolar et al. (2012) compared alternate form reliability using the same probes with a score of 0.86 and with different probes with a score of 0.74. Conoyer et al. (2017) found alternate form reliability around 0.80.

Criterion validity is investigated by examining whether two or more different types of measurements of the same skill lead to similar results (Jenkins & Jewell, 1993). However, maze and ORF tests have the same test format, while achievement tests use different types of measurements for the same skill between different grades (Jenkins & Jewell, 1993). In addition, achievement tests that are geared towards measuring the performance for high-stakes decisions, do not provide data on student progress, and are not designed for repeated administrations throughout the school year (Crawford, Tindal, & Stieber, 2001).

The frequent correlation of CBM results with achievement tests provides teachers with the data they need to plan interventions and monitor the progress of students, that achievement tests did not provide (Crawford et al., 2001). However, in order for the correlation between the two tests to be considered reliable, the sample size, the proportion ELLs, the proportion of students with learning disabilities and the time between administrations of CBM and state achievement tests should be taken into account (Yeo, 2010). Differences in maze with different selection criteria of distractors and different scoring types (correct answers and correct minus wrong answers), did not seem to affect degrees of reliability (Conoyer et al., 2017).

Measuring growth seems to be one of the most challenging and complex issues for CBM researchers and developers. Technical features of slope are difficult to be examined because researchers usually focused on technical features of static score, as they are more familiar with investigating traditional psychometric characteristics (Fuchs, 2004). Numerous terms are included in technical features of slope, the most common of which are mean growth rate, reliability of slope, Standard Error of Estimate (SEE), and SEE of mean growth rate.

Shin et al., (2000) examined the mean growth rate in maze scores, of general education students and remedial education students, recording 1.20 more correct answers per month for the first group and 0.91 for the second. Although the difference seems large, it did not prove to be statistically significant. The researchers suggest that one possible reason may have been the successful remedial education services that students received (Shin et al., 2000). On the other hand, Chung et al. (2018), found that growth rates for 7th grade students, with and without learning difficulties, did not differ significantly, while test scores differed.

SEE is important, because the higher it is, the less useful CBM test results will be for monitoring progress (Fuchs & Fuchs, 1992). Fuchs and Fuchs (1992) reported that the average progress per week was 0.39 words and the SEE 3.09. Monthly progress (1.07 words) recorded by Shin et al. (2000) was statistically significant, and its degree of reliability was 0.66, indicating that 66% of the rec-

orded progress could be attributed to a real change in student performance. SEE of mean growth rate was close to the findings of Fuchs and Fuchs (1992).

Investigating the terms of the sensitivity of the mean growth rate, Shin et al. (2000) recorded statistically significant progress that reveals significant intraindividual differences between measurements. On the other hand, Chung et al. (2018), compared the achievement and progress of 7th grade students with and without learning difficulties, and found significant differences in test scores, but not in growth rates.

Students' recorded progress in reading using the maze appears to be consistent and reliable (approximately 0.4 words per week) (Fuchs & Fuchs, 1992; Shin et al., 2000). However, it is not as useful as reading fluency which is much larger quantitatively (approximately 1 word per week) (Christ et al., 2010). This quantitative disadvantage is argued to have the potential to be offset by converting the maze grades to ORF grades (Fuchs & Fuchs, 1992).

More specifically, Jenkins and Jewell (1993), investigating the degree of correlation of ORF with achievement tests across grade levels, found correlations around 0.85 for grades 2 - 4 which reached 0.63 for grade 6. Corresponding fluctuations were not observed for the maze, which recorded correlations from 0.63 to 0.76. However, maze correlations with achievement tests had greater stability across grades, while lower correlations were observed for the first grades of school.

As shown above, the degree of correlation of the maze fluency with standardized reading achievement tests is not as high as reading ability, but appears more stable over time in all grades of Elementary and High School, in contrast to reading ability which is higher for smaller grades (1 - 4) and decreases for older students (Jenkins & Jewell, 1993; Wayman et al., 2007).

A recent systematic literature review by Shin and McMaster (2019), investigating correlation coefficients for reading fluency and the maze, with achievement test for reading comprehension in students from 1st - 10th grade, contradict previous findings. Their findings suggested that reading comprehension may be a better predictor of reading comprehension than the maze. However, the differences were not large, indicating that both types of measurements can be useful predictors of reading comprehension.

The studies presented thus far, provide evidence that the technical weaknesses that the maze seems to present as a measurement tool for monitoring reading progress in the context of CBM versus the ORF measurement, are compensated by its advantages for simultaneously administration in whole classes and automatically scoring (Shin et al., 2000).

8. The Maze for Universal Screening

The administration of maze fluency test can be done in cases of students who are at risk of developing difficulties in reading comprehension, and to have a complementary feasibility with the measurement of reading fluency in order to con-

tribute in a more complete reading profile of the students (Shinn & Shinn, 2002a). Of course, in the above case, doubts are expressed as to whether the maze provides useful information in combination with other measurements (Ardoin et al., 2004). However, to be used as a screening tool, it would be better to extend the administration time or administer two maze texts, as it has been found that the longer the administration time is, the greater the reliability coefficients are (Chung et al., 2018; Conoyer et al., 2017).

The maze can be used in the context of universal screening to identify students who have difficulty with lower reading comprehension skills and can be improved through revisions to focus more on comprehension skills (Muijselaar, Kendeou, de Jong, & van den Broek, 2017). Finally, it has great predictive validity for scoring in achievement tests, even in 8th grade students (Conoyer et al., 2017). All the above, combined with the maze's ability to be administered simultaneously to whole classes and automatically score, render it an important and useful tool for the purposes of universal screening assessment.

9. Which Components of Reading Does the Maze Measure?

Reading is a complex process involving several subskills, such as word decoding, reading fluency, the ability to deduce conclusions and retain information in memory, having as ultimate goal of reading comprehension (Jenkins et al., 2003; National Reading Panel & National Institute of Child Health and Human Development, 2000; Oakhill & Cain, 2012). There is no consensus among researchers on skills most related to reading comprehension. The "simple" theories regarding reading claim that it is the sum of decoding with language comprehension (Gough & Tunmer, 1986), while others place more emphasis on the quality of lexical representations (Perfetti, 2007).

The theoretical background, type of tests, and materials used in different reading assessment tools, can lead to different results and even measurements of the same skills (Cutting & Scarborough, 2006). Research efforts to examine maze criterion and content validity have been conducted primarily with achievement tests (Chung et al., 2018; Conoyer et al., 2017; Guthrie et al., 1974; Jenkins & Jewell, 1992, 1993; Muijselaar et al., 2017; Wiley & Deno, 2005; Shin et al., 2000) with different maze formats in terms of test duration, distractors selection method, scoring methods and administration at different ages (Tolar et al., 2012). Therefore, correlations between maze and achievement tests are likely to differ across studies because of the differences between the tests.

Although maze fluency is used as a reading comprehension measure, the skills associated with it have not been sufficiently explored (Kendeou & Papadopoulos, 2012). It is argued that it interrelates with code-related abilities, such as reading fluency and word decoding, and less with language comprehension skills, such as vocabulary and language comprehension (Muijselaar et al., 2017). Still, this correlation remains robust over time and does not appear to be affected between grades (4 - 7 - 9 grades). In a longitudinal study with Greek-speaking students

from 1st through 2nd grade, it was found that maze performance can be predicted by rapid naming (RAN) and word reading fluency (Kendeou & Papadopoulos, 2012). Furthermore, it interrelates with vocabulary and more specifically with the breadth and efficiency of lexical representations leading to fluent reading (Kendeou, Papadopoulos, & Spanoudis, 2012).

Conoyer et al. (2017) hypothesized that different distractors selection criteria would lead to the measurement of different reading skills. That is, "difficult" distractors are more likely to fail to measure a "deeper" level of comprehension, as originally assumed by Parker et al. (1992). Students' pauses while trying to find the right answer, burden the function of working memory and make it difficult to create mental representations that lead to a "deeper" understanding of the text. On the other hand, if "easy" distractors are selected, which do not require students to make cognitive efforts to select them, as originally designed by Deno et al. (1989) in BASS, then the maze will be closer to measuring reading fluency skills (Conoyer et al., 2017). Considering the different effects coming from the choice of distractors, Conoyer et al. (2017) investigated the criterion validity of the two maze types with two standardized achievement tests and found no statistically significant differences in correlation scores. The maze, however, with the "easy" distractors seemed to be more related to the general reading ability, while those with the "difficult" ones were more related to the reading comprehension. Of course, the above need further investigation and testing in younger school students.

10. Material Selection for Conducting Maze Measures

The source of materials for R-CBM measurements has been a field of interest since the early 1990s and has been explored more in terms of ORF, less in the maze and even lesser in word reading fluency (Wayman et al., 2007). The process of creating or selecting materials for the maze has no differences from ORF, except that the texts are usually longer due to the longer maze administation time.

The initial purpose of CBM for reading was to evaluate student performance based on students' curriculum, in order the source of materials to be from each grade materials (Deno & Fuchs, 1987). This tactic resulted in unstable sources of materials, as teachers chose the texts they would use each time (Deno, 1985).

The unstable sources of materials seemed to negatively affect the technical adequacy of the measurements and led to different performance of the students (Wayman et al., 2007). For example, it has been found that in texts created for assessment purposes and tested for their level of difficulty with readability formulas, students had higher score than literature-based texts (Brown-Chidsey, Johnson, & Fernstrom, 2005).

In addition, it is important to ensure the equivalence of passages by having similar difficulty levels in the texts of each probe, as measurements become more "sensitive" in recording the student's progress (Fuchs & Shinn, 1989). Of course,

it is less important for monitoring students' progress and for evaluating the choice of teaching strategies at the classroom level, as it is enough to compare classmates' performance without referring to a norm or criterion (Fuchs & Shinn, 1989).

The sources of material directly from the curriculum gave high face validity, but more measurement errors, due to the unstable variation of text difficulty and the possible familiarity of the student with the text (Fuchs & Deno, 1994). It also proved to be time consuming for teachers to design and correct a CBM probe each time (Shinn & Shinn, 2002b).

The continuous research concluded that it was preferable to use curriculum independent texts, so that the measurements have higher validity and reliability, reduce the SEE and have higher possibility of generalizing the results between the various curricula that students can be taught (Wayman et al., 2007).

11. How Often Is the Maze Used for Measuring Growth in Reading?

Guthrie et al. (1974) used the maze for progress monitoring in reading comprehension, administering it weekly with texts of similar difficulty, in order to monitor students' progress over a period of approximately 5 weeks. Fuchs and Fuchs (1992) administered it twice a week for 18 weeks to determine the progress of 63 students who had been diagnosed with reading difficulties with an average age of 12 years. Shinn and Shinn (2002b) at Aimsweb, have designed it to be administered as a benchmark only three times a year, and Deno et al. (1989) in the context of screening assessment. In the latest edition of DIBELS 8th (University of Oregon, 2020) the maze is used as a benchmark three times a year and, regarding progress monitoring, it is recommended to be administered no more than once a month.

12. Is the Maze a Time-Consuming Method?

In the research of Guthrie et al. (1974), cited as the first "official" literature reference suggesting the maze as a progress monitoring method, administration had no time limit and students had as much time needed to complete the test. In all subsequent uses of the CBM maze text, administration time varies and may be 1 minute (Deno et al., 1989; Jenkins & Jewell, 1993; Wiley & Deno, 2005), 2 minutes (Brown-Chidsey, Johnson, & Fernstrom, 2005), 2.5 minutes (Fuchs & Fuchs, 1992), 3 minutes (Ardoin et al., 2004; University of Oregon, 2020; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Shin et al., 2000; Shinn & Shinn, 2002b) up to 10 minutes (Brown-Chidsey, Davis, & Maya, 2003).

The comparison of maze to ORF as benchmark assessments administered three times a year, in the context of universal screening, seems to disserve maze in regarding time consuming issues. The maze is usually considered a 3-minute measure and ORF as 1-minute. In fact, it would take about 5 minutes, including instruction, scoring, and data representation, to evaluate an entire class of 25

students with a maze that includes administration and scoring software. If ORF measurement is chosen, it will take about 5 minutes for each student.

13. Do Teachers and Students Prefer the Maze as a Measurement Method?

The process of administering the maze seemed to meet the teachers' requirements, as it reflects various dimensions of reading such as decoding, comprehension and fluency (Fuchs & Fuchs, 1992). It seemed a pleasant process for students compared to other assessment methods (Fuchs & Fuchs, 1992). One possible interpretation is that the students had not participated in a similar assessment process in the past, the group nature of the assessment, the use of computer, the opportunity for multiple answers without much cognitive effort and the avoidance of exposing students to aloud reading that can cause embarrassment and feelings of shame, especially in students with reading difficulties. Also, it seems that monitoring progress using the maze gives students feedback on their results, motivating them to become more responsible for their progress in reading, put more effort in the task, and attribute their success to their personal effort (Davis, Fuchs, & Fuchs, 1995).

14. Conclusion

The purpose of the current article was a literature review of CBM maze. The investigation of CBM maze showed that, with the corroboration of technology, it can be used as a solely automatic measure for administering, scoring, and representing the data from the assessment. Furthermore, its applicability for administration in whole classes can save time and resources redirecting them for diagnostic, formative, and summative assessment.

The maze initial purposes, as formulated by Guthrie et al. (1974), to render it as a reliable and valid GOM measure for universal screening and progress monitoring have proven by many studies (Chung et al., 2018; Conoyer et al., 2017; Crawford et al., 2001; Fuchs & Fuchs, 1992; Guthrie et al., 1974; Jenkins & Jewell, 1992; Jenkins & Jewell, 1993; Kendeou & Papadopoulos, 2012; Kendeou et al., 2012; Muijselaar et al., 2017; Shin et al., 2000; Shinn & Shinn, 2002a; Shin & McMaster, 2019; Tolar et al., 2012; Yeo, 2010; Wayman et al., 2007). One source of weakness, regarding compares with ORF and maze, is that the maze has much less investigated. In the meta-analysis of Shin and McMaster (2019), only 29 relevant surveys were identified for the maze fluency versus 123 for ORF. Furthermore, most of these studies focused on students in lower grades (1 - 3 grades), where the emphasis is on reading skills, such as decoding and fluency, rather than on intermediate grades, where the emphasis is on comprehension skills (Shin & McMaster, 2019).

In the maze history, as a measurement for reading, five important stages of development have transpired, which can be seen through commercial CBM tests and research papers. Several changes have occurred in many issues, such as ad-

ministration time, criteria for selection of distractors, scoring procedures, issues for representing students' growth, and sources of material, gradually improving maze utility as a measure for universal screening and progress monitoring.

The additional changes described above enhanced its instructional utility as a GOM for reading. Fuchs and Fuchs (1992) with the addition of technology, took it one step further by enabling the gradual supplanting of paper-based tests. One of the more significant findings to emerge from this study was that the maze converted into an automated procedure for administration, scoring, representing the data from the assessment, and giving educators and parents significant feedback.

This review has disclosed various questions in need of further investigation. Eventhough, the maze has almost fifty years history in the bibliography, many issues need further investigation, such as the components of reading that the maze measures (Kendeou & Papadopoulos, 2012; Kendeou, Papadopoulos, & Spanoudis, 2012; Muijselaar et al., 2017; Tolar et al., 2012), its instructional utility as a progress monitoring tool (Chung et al., 2018), differences in selection of distractors (Conoyer et al., 2017) and the use of the maze in different languages (Abu-Hamour, 2013; Kendeou & Papadopoulos, 2012; Kendeou, Papadopoulos, & Spanoudis, 2012).

Implications of the Study

The findings of this study have several practical implications for the maze users. The purposes of the assessment that the maze serves can be better understood by following maze historical development as a measurement for reading, the various modifications and additions made over time, as well as recognizing the original purpose of developing the test. A more detailed knowledge of the technical characteristics of the measurement, such as validity, reliability, mean growth rate, SEEs and SEMs, can guide teachers in the educational decision-making process, so that they may or may not collect more data for universal screening or progress monitoring purposes. Also, the understanding of reading skills measured by the maze can lead teachers to acknowledge more about their students' strengths and weaknesses in reading, to create more targeted interventions and enable them to evaluate the effectiveness of their reading instruction.

Funding

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abu-Hamour, B. (2013). The Use of the Arabic CBM Maze among Three Levels of Achievers in Jordan. *International Journal of Special Education, 28,* 150-159.

https://files.eric.ed.gov/fulltext/EJ1024425.pdf

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., Williams, K. L. et al. (2004). Examining the Incremental Benefits of Administering a Maze and Three versus One Curriculum-Based Measurement Reading Probes When Conducting Universal Screening. *School Psychology Review, 33,* 218-233. https://doi.org/10.1080/02796015.2004.12086244

Ball, C. R., & Christ, T. J. (2012). Supporting Valid Decision Making: Uses and Misuses of Assessment Data within the Context of RTI. *Psychology in the Schools, 49,* 231-244. https://doi.org/10.1002/pits.21592

Beale, I. L. (2005). Scaffolding and Integrated Assessment in Computer Assisted Learning (CAL) for Children with Learning Disabilities. *Australasian Journal of Educational Technology, 21,* 173-191. https://doi.org/10.14742/ajet.1332

Brown-Chidsey, R., Davis, L., & Maya, C. (2003). Sources of Variance in Curriculum-Based Measures of Silent Reading. *Psychology in the Schools, 40,* 363-377. https://doi.org/10.1002/pits.10095

Brown-Chidsey, R., Johnson Jr., P., & Fernstrom, R. (2005). Comparison of Grade-Level Controlled and Literature-Based Maze CBM Reading Passages. *School Psychology Review, 34,* 387-394. https://doi.org/10.1080/02796015.2005.12086293

Burns, M. K., Taylor, C. N., Warmbold-Brann, K. L., Preast, J. L., Hosp, J. L., & Ford, J. W. (2017). Empirical Synthesis of the Effect of Standard Error of Measurement on Decisions Made within Brief Experimental Analyses of Reading Fluency. *Psychology in the Schools, 54,* 640-654. https://doi.org/10.1002/pits.22022

Busch, T. W., & Reschly, A. L. (2007). Progress Monitoring in Reading: Using Curriculum-Based Measurement in a Response-to-Intervention Model. *Assessment for Effective Intervention, 32,* 223-230. https://doi.org/10.1177/15345084070320040401

Christ, T. J., & Ardoin, S. P. (2009). Curriculum-Based Measurement of Oral Reading: Passage Equivalence and Probe-Set Development. *Journal of School Psychology, 47,* 55-75. https://doi.org/10.1016/j.jsp.2008.09.004

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-Based Measurement of Oral Reading: An Evaluation of Growth Rates and Seasonal Effects among Students Served in General and Special Education. *School Psychology Review, 39,* 447-462. https://doi.org/10.1080/02796015.2010.12087765

Chung, S., Espin, C. A., & Stevenson, C. E. (2018). CBM Maze-Scores as Indicators of Reading Level and Growth for Seventh-Grade Students. *Reading and Writing, 31,* 627-648. https://doi.org/10.1007/s11145-017-9803-8

Conoyer, S. J., Lembke, E. S., Hosp, J. L., Espin, C. A., Hosp, M. K., & Poch, A. L. (2017). Getting More from Your Maze: Examining Differences in Distractors. *Reading & Writing Quarterly, 33,* 141-154. https://doi.org/10.1080/10573569.2016.1142913

Coyne, M. D., & Harn, B. A. (2006). Promoting Beginning Reading Success through Meaningful Assessment of Early Literacy Skills. *Psychology in the Schools, 43,* 33-43. https://doi.org/10.1002/pits.20127

Coyne, M. D., Zipoli Jr., R. P., & Ruby, M. F. (2006). Beginning Reading Instruction for Students at Risk for Reading Disabilities: What, How, and When. *Intervention in School and Clinic, 41,* 161-168. https://doi.org/10.1177/10534512060410030601

Crawford, L., Tindal, G., & Stieber, S. (2001). Using Oral Reading Rate to Predict Student Performance on Statewide Achievement Tests. *Educational Assessment, 7,* 303-323. https://doi.org/10.1207/S15326977EA0704_04

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive

Skills Can Depend on How Comprehension Is Measured. *Scientific Studies of Reading, 10,* 277-299. https://doi.org/10.1207/s1532799xssr1003_5

Davis, L., Fuchs, L., & Fuchs, D. (1995). "Will CBM Help Me Learn?" Students 1 Perception of the Benefits of Curriculum-Based Measurement. *Education and Treatment of Children, 18,* 19-32. http://www.jstor.org/stable/42899384

Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children, 52,* 219-232. https://doi.org/10.1177/001440298505200303

Deno, S. L., & Fuchs, L. S. (1987). Developing Curriculum-Based Measurement Systems for Data-Based Special Education Problem Solving. *Focus on Exceptional Children, 19,* 1-16. https://doi.org/10.17161/foec.v19i8.7497

Deno, S. L., Maruyama, G., Espin, C. A., & Cohen, C. (1989). *The Basic Academic Skills samples (BASS).* Minneapolis, MN: University of Minnesota.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying Valid Measures of Reading. *Exceptional Children, 49,* 36-47. https://doi.org/10.1177/001440298204900105

Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review, 33,* 188-192. https://doi.org/10.1080/02796015.2004.12086241

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic Distinctions between Instructionally Relevant Measurement Models. *Exceptional Children, 57,* 488-500. https://doi.org/10.1177/001440299105700603

Fuchs, L. S., & Deno, S. L. (1994). Must Instructionally Useful Performance Assessment Be Based in the Curriculum? *Exceptional Children, 61,* 15-24. https://doi.org/10.1177/001440299406100103

Fuchs, L. S., & Fuchs, D. (1992). Identifying a Measure for Monitoring Student Reading Progress. *School Psychology Review, 21,* 45-58. https://doi.org/10.1080/02796015.1992.12085594

Fuchs, L. S., & Shinn, M. R. (1989). Writing CBM IEP Objectives. In *Curriculum-Based Measurement: Assessing Special Children* (pp. 130-152). New York: The Guilford Press. https://www.researchgate.net/profile/Mark_Shinn/publication/232524468_Writing_CBM_IEP_objectives/links/56312cf108ae506cea677783/Writing-CBM-IEP-objectives.pdf

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative Evaluation of Academic Progress: How Much Growth Can We Expect? *School Psychology Review, 22,* 27-48. https://doi.org/10.1080/02796015.1993.12085636

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5,* 239-256. https://doi.org/10.1207/S1532799XSSR0503_3

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching Reading Comprehension Strategies to Students with Learning Disabilities: A Review of Research. *Review of Educational Research, 71,* 279-320. https://doi.org/10.3102/00346543071002279

Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education, 7,* 6-10. https://doi.org/10.1177/074193258600700104

Guthrie, J., Seifert, M., Burnham, N., & Caplan, R. (1974). The Maze Technique to Assess, Monitor Reading Comprehension. *The Reading Teacher, 28,* 161-168. http://www.jstor.org/stable/20193704

Hintze, J. M. (2009). Conceptual and Empirical Issues Related to Developing a Response-to-Intervention Framework. *Journal of Evidence-Based Practices for Schools, 9,*

128-147.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-Based Assessment. *Psychology in the Schools, 43,* 45-56. https://doi.org/10.1002/pits.20128

Hintze, J. M., Wells, C. S., Marcotte, A. M., & Solomon, B. G. (2018). Decision-Making Accuracy of CBM Progress-Monitoring Data. *Journal of Psychoeducational Assessment, 36,* 74-81. https://doi.org/10.1177/0734282917729263

Jenkins, J. R., & Jewell, M. (1992). An Examination of the Concurrent Validity of the Basic Academic Skills Samples (BASS). *Diagnostique, 17,* 273-288. https://doi.org/10.1177/153450849201700405

Jenkins, J. R., & Jewell, M. (1993). Examining the Validity of Two Measures for Formative Teaching: Reading Aloud and Maze. *Exceptional Children, 59,* 421-432. https://doi.org/10.1177/001440299305900505

Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal of Educational Psychology, 95,* 719. https://doi.org/10.1037/0022-0663.95.4.719

Kendeou, P., & Papadopoulos, T. C. (2012). The Use of CBM-Maze in Greek: A Closer Look at What It Measures. In C. A. Espin, K. L. McMaster, S. Rose, & M. M. Wayman (Eds.), *A Measure of Success: The Influence of Curriculum-Based Measurement on Education* (pp. 329-339). Minneapolis, MN: University of Minnesota Press. https://www.academia.edu/3902506/The_Use_of_Curriculum-Based_Measurement_Maze_in_Greek https://doi.org/10.5749/minnesota/9780816679706.003.0030

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing Demands of Reading Comprehension Tests in Young Readers. *Learning and Instruction, 22,* 354-367. https://doi.org/10.1016/j.learninstruc.2012.02.001

Leighton, J. P., & Gierl, M. J. (2011). *The Learning Sciences in Educational Assessment: The Role of Cognitive Models.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511996276

Linacre, J. M. (2000). *Computer-Adaptive Testing: A Methodology Whose Time Has Come* (No. 69, p. 58). MESA Memorandum. https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000_CAT.pdf

McConnell, S., & Wackerle-Hollman, A. (2015). Can We Measure the Transition to Reading? General Outcome Measures and Early Literacy Development from Preschool to Early Elementary Grades. *AERA Open, 2.* https://doi.org/10.1177/2332858416653756

Muijselaar, M. M., Kendeou, P., de Jong, P. F., & van den Broek, P. W. (2017). What Does the CBM-Maze Test Measure? *Scientific Studies of Reading, 21,* 120-132. https://doi.org/10.1080/10888438.2016.1263994

National Reading Panel (U.S.), & National Institute of Child Health and Human Development (U.S.) (2000). *Report of the National Reading Panel: Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups.* Washington DC: National Institute of Child Health and Human Development, National Institutes of Health.

Oakhill, J. V., & Cain, K. (2012). The Precursors of Reading Ability in Young Readers: Evidence from a Four-Year Longitudinal Study. *Scientific Studies of Reading, 16,* 91-121. https://doi.org/10.1080/10888438.2010.529219

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The Maze as a Classroom-Based Read-

ing Measure: Construction Methods, Reliability, and Validity. *The Journal of Special Education, 26,* 195-218. https://doi.org/10.1177/002246699202600205

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading, 11,* 357-383. https://doi.org/10.1080/10888430701530730

Petscher, Y., Cummings, K. D., Biancarosa, G., & Fien, H. (2013). Advanced (Measurement) Applications of Curriculum-Based Measurement in Reading. *Assessment for Effective Intervention, 38,* 71-75. https://doi.org/10.1177/1534508412461434

Protopappas, A., & Skaloubakas, Ch. (2008). Assessment of Reading Fluency for Reading Difficulties Identification. *Psychologia, 15,* 267-289. (In Greek)

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-Based Measurement Oral Reading as an Indicator of Reading Achievement: A Meta-Analysis of the Correlational Evidence. *Journal of School Psychology, 47,* 427-469. https://doi.org/10.1016/j.jsp.2009.07.001

Shin, J., & McMaster, K. (2019). Relations between CBM (Oral Reading and Maze) and Reading Comprehension on State Achievement Tests: A Meta-Analysis. *Journal of School Psychology, 73,* 131-149. https://doi.org/10.1016/j.jsp.2019.03.005

Shin, J., Deno, S. L., & Espin, C. (2000). Technical Adequacy of the Maze Task for Curriculum-Based Measurement of Reading Growth. *The Journal of Special Education, 34,* 164-172. https://doi.org/10.1177/002246690003400305

Shinn, M. M., & Shinn, M. R. (2002a). *AIMSweb Training Workbook: Administration and Scoring of Reading Curriculum-Based Measurement (R-CBM) for Use in General Outcome Measurement.* Eden Prairie, MN: Edformation. https://www.researchgate.net/publication/267793797

Shinn, M. R., & Shinn, M. M. (2002b). *AIMSweb Training Workbook: Administration and Scoring of Reading Maze for Use in General Outcome Measurement.* Eden Prairie, MN: Edformation. https://www.cnyric.org/tfiles/folder1052/Administration%20and%20Scoring%20of%20Reading%20MAZE.pdf

Shinn, M. R., Knutson, N., Good III, R. H., Tilly III, W. D., & Collins, V. L. (1992). Curriculum-Based Measurement of Oral Reading Fluency: A Confirmatory Analysis of Its Relation to Reading. *School Psychology Review, 21,* 459-479. https://doi.org/10.1080/02796015.1992.12085629

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using Progress-Monitoring Data to Improve Instructional Decision Making. *Preventing School Failure: Alternative Education for Children and Youth, 52,* 48-58. https://doi.org/10.3200/PSFL.52.2.48-58

Tolar, T. D., Barth, A. E., Francis, D. J., Fletcher, J. M., Stuebing, K. K., & Vaughn, S. (2012). Psychometric Properties of Maze Tasks in Middle School Students. *Assessment for Effective Intervention: Official Journal of the Council for Educational Diagnostic Services, 37,* 131-146. https://doi.org/10.1177/1534508411413913

Tzivinikou, S. (2018). Screening Young Children at Risk for Reading Failure. In *Early Childhood Education* (pp. 1-15). London: IntechOpen. https://doi.org/10.5772/intechopen.82081

University of Oregon (2020). *8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): Administration and Scoring Guide.* Eugene, OR: University of Oregon. https://dibels.uoregon.edu/docs/materials/d8/dibels_8_admin_and_scoring_guide_05_2020.pdf

Van Norman, E. R., & Christ, T. J. (2016). Curriculum-Based Measurement of Reading: Accuracy of Recommendations from Three-Point Decision Rules. *School Psychology*

Review, 45, 296-309. https://doi.org/10.17105/SPR45-3.296-309

Van Norman, E. R., Maki, K. E., Burns, M. K., McComas, J. J., & Helman, L. (2018). Comparison of Progress Monitoring Data from General Outcome Measures and Specific Subskill Mastery Measures for Reading. *Journal of School Psychology, 67,* 179-189. https://doi.org/10.1016/j.jsp.2018.02.002

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *The Journal of Special Education, 41,* 85-120. https://doi.org/10.1177/00224669070410020401

Wiley, H. I., & Deno, S. L. (2005). Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment. *Remedial and Special Education, 26,* 207-214. https://doi.org/10.1177/07419325050260040301

Yeo, S. (2010). Predicting Performance on State Achievement Tests Using Curriculum-Based Measurement in Reading: A Multilevel Meta-Analysis. *Remedial and Special Education, 31,* 412-422. https://doi.org/10.1177/0741932508327463