This paper describes the maze assessment and its application in accurate identification of student achievement levels in reading. It also recommends strategies for planning differentiated instruction.

Using Maze Assessment in the Classroom

Research in Education Inc.

Applying research best practices to improve the quality and effectiveness of instructional and assessment programs for K-12 learners

Email: info@researchineducationinc.com

Authored by: Edina Torlakovic and Ernest Balajthy

Table of Contents

Using Maze Assessments in the Classroom

Cloze and maze assessments are well-established, research-based tests of student reading that measure word recognition and comprehension ability. They are valuable tools for busy classrooms because they are quick, easy to administer, and provide a reliable data point for determining student reading levels. In both assessments, students are presented with a reading passage. The first sentence of the passage is intact. In the remainder of the passage, words are deleted and replaced with blank lines. Students are asked to identify words that might appropriately fit the blanks. In cloze assessments, the deletions are completely blank and students are asked to write in a word that would fit the blank appropriately. In the maze assessment, students are given a choice of three words to fill in the deletion.

Word deletion assessment as a measure of student reading ability was originally proposed by Wilson Taylor (1953), but these tests first became popular in the 1970s and 1980s, largely as a result of the work of University of Chicago researcher John Bormuth (1969). Bormuth argued that the cloze and maze procedures were richer, more direct, and highly efficient measures of the relationship between student reading performance and reading materials/levels than traditional reading tests based on multiple choice questions or on oral reading.

As cloze assessment became well-established in schools in the 1980s, Bormuth and others introduced a related measure of assessment, the maze procedure. The maze was welcomed by schools for several reasons. It is an easier task for students to make a choice from three possibilities rather than facing a blank line in a passage. The selection format is more familiar to students than the cloze, so that initial acclimatization to the procedure takes less time. Finally, the use of maze testing in computer-based assessment is more efficient than the cloze format due to its simpler design and scoring.

When properly administered, maze assessments can achieve similar results more efficiently than time-consuming informal reading inventories (IRIs), which are diagnostic reading tests that are comprised of oral reading passages, miscue analysis, and comprehension questions.

Purpose

The purpose of this paper is to provide an overview of the maze assessment and what it measures, as well as practical applications in the classroom for differentiating instruction. Maze is a simple and efficient tool for student assessment, but it can also be used in a variety of ways to enhance student learning. Teachers require assessments that will both place students at accurate instructional levels and predict student scores on federally mandated state tests. Maze assessment provides an accurate measure to meet both of these purposes.

What is the Maze?

The maze assessment requires students to read sets of passages made up of sentences that have deleted words. Traditional procedures include the following (Deno, 1985; Espin & Foegen, 1996; Muijselaar, 2017):

- Intact passages are given readability analyses by the test developer to identify their reading levels. Publishers typically identify these levels for teacher use.
- Passages are administered at the students' grade level. For example, 7th graders will be administered passages with readability at the 7th-grade level.
- Students choose the correct word to complete the sentence from three choices, two of which are distractors (also called foils).
- Maze tests can be administered on paper or by computer. In both models, each deletion is replaced by a set of three distractors. In paper-based administrations, students circle or underline their choices. In computer-based testing, students click on their choices.

Research-based Norms

Maze testing generates a percentage score based on the number of correct selections out of the total available in the passage. Results are then assigned to one of the three reading ranges: frustration, instructional and independent. Researchers vary on the percentages for each of these reading ranges. Table 1 provides a summary of published research on reading ranges.

Guthrie, et al. (1974) offered the first research establishing maze norms (Table 1). Later work was carried out by Fuchs, et al. (1993) and Jenkins and Jewell (1993). Norms developed under early federally funded research grants to establish the DIBELS assessment system are also available.

Table 1: Maze Assessment Norms

Research Reference	Frustration Range	Instructional Range	Independent Range
Guthrie, et al., 1974	Below 60%	60—84%	85—100%
Degrees of Reading Power (cited in Parker, Hasbrouck, & Tindal, 1992)	Approximately 50%	Approximately 75%	90%—above
Feely, 1975	75% or less	80%—90%	90% and above
Harris & Sipay, 1985	Below 90%	90% and above	

Number of passages

The accuracy of the maze assessment is well-established. A general approach to increasing the accuracy (that is, statistical reliability) of assessments is to make them longer. Chung, Espin, and Stevenson (2018) found this to be true of maze assessments.

In addition to passage length, the number of passages also affects reliability. Use of multiple passages is common in research studies on the maze, where accuracy plays a key role. Muijselaar (2017), for example, provided students with three passages, as did Espin, et al. (2010). Use of multiple passages increases the length and accuracy of the assessment, and it also provides accountability for effects of passage topic and discipline area. Shinn (2017) reported that most research studies on the maze assessment use three passages.

Score calculation across multiple forms

Maze scores are usually derived from a raw score, which is the simple number of correct items. Current research does not identify any one best approach to scoring. Conoyer, et al. (2017), for example, surveyed research to conclude that few studies have addressed the comparative impact of different scoring methods on maze test quality.

Muijselaar, et al. (2017) followed a common policy of calculating an average adjusted score across three forms. The result can be input into a table of results designed for a single maze administration.

In maze assessment using multiple forms, the upper and lower scores can be discarded and only the middle score is counted (e.g., Wright, 2013). This policy, easy for hand-scoring as it involves no calculations, attempts to account for scores that, for whatever reason, are outliers.

Practical Ideas for Using Maze Assessment to Differentiate Instruction

Using maze for placement at approximate levels of instruction

Maze is commonly used for identification of students' *instructional reading level*. A key purpose of this identification is to group children for instruction. A student's instructional level is the level at which instruction and learning are carried out with the most effectiveness. It is a level that is challenging for the student, but not so difficult as to be discouraging or to make successful learning overly difficult to achieve.

Computer-based maze tests are automatically scored to provide a percentage score and assignment to a corresponding frustration, instructional and independent level. Publishers of paper-based maze assessments usually include scoring and analysis charts designed specifically for their own products.

If the raw score indicates that the passage was at the student's instructional level, the child is guided into reading instruction at that reading level. For example, consider a 4th-grade student who takes a maze assessment based on a passage at a 4th-grade reading level. If the student's raw score is identified as being at the instructional level, the student may be guided into instruction geared toward 4th-grade difficulty of text, reading standards and reading objectives.

If the raw score indicates that the passage is above or below the student's instructional level, this suggests that the student's reading level is either higher or lower than grade level. The publisher may suggest using instructional materials above or below the student's actual grade.

For example, a 4th-grade student may obtain a raw score on a 4th-grade passage that is higher than the instructional range, indicating independent reading at 4th grade. The student may then be guided into instruction at the 5th- or 6th-grade level.

Another 4th-grade student may obtain a raw score on a 4th-grade maze passage that is lower than the instructional level range, indicating frustration in reading at that level. The student may then be guided into instruction at the 3rd- or 2nd-grade level.

Using maze for precision in student placement at instructional level

Some teachers may wish to take the maze assessment a step further to pinpoint the exact instructional reading level of students. This process is often called benchmarking, especially if the initial score the student receives is to be compared to later scores.

The teacher's goal in this process is to use multiple passages at different reading difficulty levels to zero in on the student's precise instructional level – the same method as used in informal reading inventories.

First, consider the student who struggled with the initial maze reading task and scored below instructional level on a 4th-grade passage. In order to zero in on this student's precise instructional level, the teacher now proceeds to administer a maze assessment at a difficulty level one grade level *below* the student's actual grade. (In our example, this second maze passage would be at the 3rd-grade level.) If the student scores in the instructional range on this passage, the process stops—we have identified the child's instructional level (in our example, the third grade reading level). This precise instructional level should be comparable to results from standardized reading tests and state reading tests.

In the next example, the 4th-grader who scored above instructional level would now be administered a maze assessment at a difficulty level one grade level *above* the student's actual grade (in our example, at the 5th-grade level). If the student scores in the instructional range on this passage, the process continues. The next higher level is then administered (6th grade). The process continues up through the grades until the student scores in the *frustration* range at a tested level. At that point, the assessment process ends, as the test administrator will identify the *highest instructional grade level* as the student's precise instructional reading level. (For example, if the student scored in the instructional range in the 6th and 7th grade maze texts but scored frustration in the 8th grade, the precise instructional level would be 7th grade.)

Lastly, the student who scored instructional at the 4th-grade level should be tested at the 4th grade. As with the example above, it is possible that the student's precise *highest* instructional level may be above 4th grade. The test should continue up through the grades to find the highest instructional level.

Using maze to predict results of state tests or other standardized tests

An important use of maze assessment is to predict student performance on state tests or other standardized tests. In particular, educators want to identify students who are at risk of poor performance on high-stakes assessments, in order to provide appropriate interventions.

The process of using maze for this purpose is described in the section titled "Using maze for precision in student placement at instructional level." By using the procedures described there, a teacher can zero in on the student's precise instructional level (that is, the highest grade level at which a student scores in the instructional level range on the maze assessment). This level is designed to correspond to the reading level reported on standardized testing.

Using maze for progress monitoring

The maze assessment is a well-recognized tool for the long-standing philosophy that emphasizes the importance of obtaining an accurate measure of students' current reading performance.

Beginning with the Response to Intervention (RTI) movement (also referred to as Multi-Tiered Systems of Support—MTSS) in the early 2000's, many schools began to focus on Curriculum-Based Measurement (CBM). It highlights the importance of continuous progress monitoring of student achievement. Rather than assessing students on a once-a-year basis, student achievement (especially of struggling students receiving instructional interventions) was monitored as often as once each week. The frequency of these progress monitoring assessments was made possible by use of very short tests (often called CBMs or fluency tests or, simply, progress monitoring).

The earliest types of reading progress monitoring assessments were based on students' oral reading, especially for first graders. As schools became interested in assessing older students, researchers looked for instruments that would offer a stronger reading comprehension component than provided by the oral reading assessments. Since then, maze assessments have played a key role in progress monitoring efforts.

The emphasis in the use of the maze for progress monitoring is less on identifying a particular instructional level (as described in earlier sections of this paper) and more on providing short, comparable tests. Maze CBMs are typically based on short passages written at students' reading level (that is, the level at which the individual student reads, not the student's actual grade level in school). Students are given three minutes to answer as many maze items as possible in the

passage. The final score is usually the number correct (raw score) at the end of the three minute period. This score is then charted and compared to scores in ensuing weeks and months. A charted trend line of raw scores that increases at a desired rate indicates the ongoing success of the intervention. A trend line that fails to increase as desired indicates that a change in intervention is necessary.

Using a scoring system that accounts for guessing

Random guessing can skew test results, giving students an artificially high score. There is no clear body of research evidence that validates scoring systems that account for random guessing, but common sense suggests that on occasion teachers might be confronted with such problems.

Some maze researchers do not account for guessing and use simple number-correct scores as the final score (Wright, 2013). Muijselaar, et al. (2017) and Conoyer, et al. (2017) both calculated a final adjusted score by subtracting the number of incorrect responses from the number of correct responses, a common—but far from universal—procedure among current researchers.

Chung, Espin, and Stevenson (2018) chose to use another approach that may be useful for timed maze administrations. Their final scores were not adjusted; the number correct was the final score. They identified potential guessers by the combined number of correct and incorrect responses. Students' scores were identified as invalid if they produced a larger than expected combined number of responses. These researchers defined "larger than expected" as greatly above the mean correct/incorrect group score for each passage.

Provide acculturative experiences with maze prior to the test

Maze procedure is not a common instructional method. When first confronted by the task, students may perform much more poorly than they will once they are more familiar with it. This can result in initial poor performance due to the nature of the maze task, *not* to the students' level of reading achievement.

If initial student placement is carried out on the basis of these low scores, the placement may be flawed. Rapid maze gains in the weeks immediately following students' first experience with the task will be based on gaining familiarity with maze, *not* with actual growth in reading ability. In sum, prior familiarity with maze is useful for initial stability of scores.

Wright (2013) provides a very brief practice exercise in his instructions for use of maze. A somewhat longer initial pre-test practice period would seem advisable if teachers want to increase the accuracy of the assessment system.

Space multiple passage assessments over more than one day

When using multiple passages for a maze assessment, in order to increase accuracy even more, administrations can be paced over a period of days rather than at one time. This is recommended in order to account for any outside influences on a child's performance on a particular day. This sensible policy has been suggested (e.g., Bradley, Ackerson, & Ames, 1978), but untested in research. Another approach is to examine test results and re-test any students whose patterns of errors show lack of focus or understanding of the activity.

Carefully examine passages for equivalence of difficulty

The difficulty of maze passages is usually analyzed by use of readability measures, but no readability measure can account for all factors relating to the challenges presented by a piece of text. As a result, test passages can have exactly identical readability scores but be considerably different in terms of their actual reading difficulty.

Account for struggling readers in the assessment

As noted in sections above, maze tests used for benchmarking are usually carried out at the students' grade level. That is, all 7th-grade students receive maze passages with readability at the 7th-grade level.

In progress monitoring, maze tests are often carried out at the students' instructional level. For example, 7th-graders reading at the 5th-grade level are given intervention instruction at the 5th grade and are tested regularly during the school year using maze passages at the 5th-grade level.

This appropriate modification of assessment to meet the needs of struggling students also has implications for maze placement testing. Seventh graders who read at the 5th-grade level will struggle and perhaps give up on a maze test at the 7th grade, their *frustration level.* In fact, about one-third of 7th-graders will find a 7th-grade passage to be at their frustration level. Giving up on a test will result in an inaccurate result and might place the student well below his or her instructional level due to an early stoppage of effort.

To mitigate these effects, a variety of possibilities can be explored. Teachers can administer a follow-up maze test at a lower level to any students whose results indicated frustration at the tested level. Another possibility would be to have the teacher predetermine, from observation and past assessment, what would be an appropriate level at which to administer the initial maze assessment. Still yet another approach, though more complex in terms of test design, would be to test at multiple levels in a single test administration.

Factor additional measures into students' level placement

In determining a student's achievement level, the principle of using multiple measures is a cardinal rule of assessment. That is, high-stakes decisions such as what level of reading instruction to give a student are best based on multiple measures.

Rather than basing a student's placement solely on the results of a maze assessment, additional measures such as previous year's test scores, teacher observations and other formal or informal reading tests can also be used in determining the students' achievement level. Computer-based programs using the maze assessment should allow for manual override of placement levels to give teachers more control over the instructional environment.

Summary

The statistical reliability and validity of the maze assessment has been established through decades of research. It provides teachers with a simple, accurate measurement tool that functions in two ways to improve classroom instruction. First, the maze identifies students' reading ability levels, both for those whose comprehension development is satisfactory and for a broad range of struggling readers. Second, it monitors the progress of students during instruction and intervention. Online maze assessment simplifies and expedites this identification and progress monitoring and can be an effective tool in research-based instruction.

References

Bormuth, J. R. (1969). Factor validity of cloze tests as measures of reading comprehension. *Reading Research Quarterly, 4,* 358-365.

Bradley, J. M., Ackerson, G., & Ames, W. S. (1978). The reliability of the maze procedure. *Journal of Reading Behavior, 10,* 291-296.

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: An Interdisciplinary Journal, 31,* 627-648.

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.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 16*, 99-104.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 61*, 497-514.

Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research and Practice, 25,* 60-75.

Feely, T.M. (1975). How to match reading materials to student reading levels: II. The cloze and the maze. *Social Studies, 66,* 252-258.

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.

Guthrie, J. T., Seifert, M., Burnham, N. A., & Caplan, R. I. (1974). The maze technique to assess, monitor reading comprehension. *The Reading Teacher*, *28,* 161-168.

Harris, A. J., & Sipay, E. R. (1985). *How to increase reading ability: A guide to developmental and remedial methods*. New York: Longman.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59,* 421-432.

Muijselaar, M. M. L., 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.

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction, reliability, and validity. *Journal of Special Education, 26*, 195-218.

Shinn, J. (2017). *Relations between CBM (oral reading and maze) and reading comprehension on state achievement tests: A meta-analysis.* Dissertation, University of Minnesota.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30,* 415-433.

Wright, J. (2013). How To: Assess Reading Comprehension with CBM Maze Passages. www.interventioncentral.org

For more information about this publication please contact:

Edina Torlakovic, PhD (ABD),
Director of Educational Program Design, Development and Evaluation
Research in Education Inc.
edina_@researchineducationinc.com