TECHNICAL REPORT #1:

Seamless and Flexible Progress Monitoring: Age and Skill Level Extensions in Reading

Christine A. Espin, Teri Wallace, Miya Miura Wayman, Renata Ticha, Hilda Ives Wiley, and Xiaoqing Du

RIPM Year 2: 2004 – 2005

Date of Study: September 2004 – June 2005

September 2009

Note: Data set and data collection procedures are the same as Technical Report #2 and #10.

Produced by the Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

Seamless and Flexible Progress Monitoring:
Age and Skill Level Extensions in Reading

Curriculum-based Measurement (CBM) was developed in the late 1970s as a method for teachers to monitor student progress and evaluate instructional programs (see Deno, 1985). In recent years, there has been a great deal of research on CBM, especially in the area of reading (see reviews by Marston, 1989; Wayman, Wallace, Wiley, Ticha & Espin, 2007). Although across studies various age and skill levels have been addressed, most of the research has focused on growth within a single year for students within a particular age or skill level groups. However, the research base is now so extensive on the development of CBM measures in reading, that Wayman et al. (2007) suggest that the time may be right to consider the development of a seamless and flexible system of progress monitoring that can span age and skill levels.

What exactly is meant by a seamless and flexible system of progress monitoring? Wallace, Espin, McMaster, Deno, and Foegen (2007) define a seamless and flexible system of progress monitoring as one that can be used across students of different age and skill levels in different settings and different curricula. Whereas in a typical CBM approach, student progress is measured *within* a school year, in a seamless and flexible approach, student progress is measured *across* school years. This type of progress monitoring is illustrated in Figure 1 in which growth from grade K to 12 is represented for two students: Mike, an average performing student, and Kelly, a lower performing student. Three different but connected lines are used to represent growth for each student. These lines represent the different measurement procedures used for Mike and Kelly at that time in their school career. For example, initial progress might have been monitored with a word identification measure, but later progress might have been monitored with a reading aloud measure. Note that even though the specific measure changes, performance on

the various measures is linked to create a picture of growth over 12 years of school. Normative growth rates across grades K through 12 are also represented in Figure 1 via the solid diagonal lines. Thus, Mike's and Kelly's growth rates are compared not only to each other but also to students at a similar level of performance.

Figure 1 is a hypothetical example. What would be needed to actually produce a graph such as that presented in Figure 1? First, it would be necessary to determine which measures were technically adequate for monitoring progress at various age and skill levels. Although it is possible that the same measure might be used to monitor student progress from grades K to 12, it is also possible that the measure might need to change as students became older or more skilled. Second, it would be necessary to determine a method for linking different measures to produce a picture of growth such as that presented in Figure 1. Finally, it would be necessary to develop a set of standard materials that could be used consistently across time and across students if one were to want to compare growth rates of individual students to normative growth rates of peers.

In this study, we address the first step described above in the move toward the development of a seamless and flexible system of progress monitoring: We address the question of whether the same measure can be used as an indicator of reading performance at various age and skill levels, or whether that measure needs to change with age and skill level. Note that although we have used the terms "age" and "skill" levels synonymously to this point, we address the two issues separately in the research. That is to say, although we examine characteristics of measures across various age levels, we also examine characteristics across skill levels within one age range. Before describing the study, we present a brief overview of the research on CBM reading.

Existing CBM Reading Research

The early CBM research in reading focused on the development of measures at the elementary-school level. This research supported the use of a 1-minute reading aloud measure as an indicator of general reading performance. Words read correctly (WRC) in one minute correlated with other measures of reading performance and reflected growth over time (see Deno, 1985; Marston, 1989 and Wayman et al., 2007 for reviews of this research). Further, WRC in one minute was found to be a good indicator of reading comprehension, relating to both formal and informal measures of reading comprehension (L. S. Fuchs, Fuchs, & Maxwell, 1988).

Later research extended the development of CBM measures to the secondary-school level. This research initially focused on the development of measures of content-area reading and learning (e.g., see Espin & Tindal, 1998, for a review), but later shifted to the development of measures of generalized reading proficiency (e.g., Espin, Wallace, Lembke, Campbell, & Long, 2009; Ticha, Espin, & Wayman, 2009). This later research suggested that the nature of the measures used to monitor progress in reading might need to change with the age of the student. In both Espin et al. (2009) and Ticha et al. (2009), despite the fact that both reading aloud and maze selection were strongly related to performance on reading criterion measures, only maze selection reflected substantial growth rates over time and only the growth rates for maze selection were related to performance and change on the criterion measures.

The possibility that the CBM measures used to monitor reading progress might need to change as students become older had been noted in earlier research. Jenkins and Jewell (1993) examined the validity of CBM reading aloud and maze selection measures for students in grades 2 through 6. Their results revealed a declining trend in correlations for reading aloud from grades 2 to 6, but consistent correlations for maze selection across the grades. Yovanoff, Duesbery, Alonzo, and Tindal (2005) compared the relative importance of vocabulary and reading fluency

as measurement dimensions of reading comprehension for students in grade 4 through 8. Results revealed that, whereas vocabulary knowledge consistently predicted reading comprehension across the grades, the effects of reading fluency decreased (although remained important) after grade 5. MacMillan (2000) examined yearly growth on reading aloud measures for students in grades 2 through 7, and found that, although growth was seen in all grades, the magnitude of that growth decreased with grade.

In sum, results of individual studies conducted at different age levels and results of cross-age studies suggest that the nature of CBM reading measures might need to change with student age and skill level. However, most of the previous research was conducted within a level (e.g., elementary- or middle-school level), or spanned only two levels (e.g., elementary- to middle-school). To date, no studies have been conducted that include participants from elementary to high school, and in fact, very few studies have included high school students at all in their samples. Further, much of the previous research in CBM reading has focused on the use of a reading aloud measure. Fewer studies have examined the technical adequacy of the maze-selection measure.

The purpose of our study was to compare the technical adequacy of two CBM reading measures across students of various skill levels at elementary-, middle-, and high-school levels. In examining technical adequacy, we considered not only the typical characteristics desirable for CBM measures as outlined in Deno (1985; i.e., reliability, validity, efficiency, sensitivity to growth), but also the characteristic of *durability* or the extent to which a measure maintains its reliability and validity across a range of age and skill levels (Wallace et al. 2007). Durability is a characteristic specific to the purpose of developing a seamless and flexible system of progress monitoring. Thus, all other things being equal, a measure that is valid and reliable for monitoring

growth across grades 3 to 12 is more desirable than a measure that is valid and reliable only across grades 3 to 6.

We focus our study on students in grades 3, 5, 8 and 10. We include a range of skill levels at each grade in order to examine characteristics of the measures for relatively lower- and higher-performing students within each grade level. We compare two CBM reading measures that had been found in previous research to be appropriate for students across these grade levels – reading aloud and maze selection. In addition to examining different measures, we examine variations in the duration and difficulty level of the measures. With respect to sample duration, in reading aloud, we examine a 1-minute sample because previous research has supported the validity and reliability of a 1-minute sample at both the elementary secondary-level (see Deno, 1985; Marston, 1989; Wayman et al., 2007; Espin et al., 2009; Ticha et al., 2009). For maze selection, previous research had revealed a potential increase in technical adequacy with an increase in time at the secondary-school level (Espin et al., 2009; Ticha et al., 2009); thus, we compared 1-, 2- and 3-minute samples for maze selection. With respect to difficulty level, previous research had demonstrated flexibility of CBM measures with respect to difficulty level (see Wayman et al., 2007, for a review of this literature), but most of that research had been conducted within grade levels. In this study, we compared the effects of difficulty level across grade levels. Specifically, we examined differences in technical characteristics for grade-level and common-level passages. Grade-level passages are passages approximately at the students' grade level; common-level passages are passages at an approximate 4th-grade level.

For reliability of the measures, we examine alternate-form reliability, an important form of reliability for repeated measurement purposes. For validity, we examine both concurrent and predictive validity. We use as our criterion measures scores on a standardized achievement test

and on state standards tests. For sensitivity to growth, we examine changes in scores on the measures from Fall to Spring for students within each grade, and we examine changes in scores across grades 3 to 10.

Four research questions were addressed in our research:

(a) What are the reliability, validity, and sensitivity to growth of CBM reading measures as indicators of general reading proficiency?

(b) Do reliability, validity, and sensitivity to growth differ with type of measure, sample duration, or difficulty level?

(c) Do reliability, validity, and sensitivity to growth differ with student grade level?

(d) Do reliability, validity, and sensitivity to growth differ with student skill levels within grade?

Method

Participants

Participants were 111 3rd graders (56% female and 44% male), 130 5th graders (48% female and 52% male), 90 8th graders (61% female and 39% male), and 178 10th graders (51% female and 49% male) from one urban and one rural Midwestern school district (see Table 1). The urban district had an enrollment of 40,499. Sixty-eight percent of students qualified for free/reduced lunch, 15% received special education services and 23% received English Language Learner services. The demographic make-up of the urban district was 4% American Indian, 12% Asian Pacific American, 42% African or African American, 14% Hispanic/Latino/Chicano, and 27% White. The rural district had an enrollment of 3,540. Seventeen percent of students qualified for free/reduced lunch, 10% received special education services and 1% received

English Language Learner services. The demographic make-up of the rural district was 1% American Indian, 2% Asian Pacific American, 1% African or African American, 1% Hispanic/Latino/Chicano, and 95% White.[1]

Participants from the urban district were from one K-8 school and one high school. Participants from the rural district were from two K-5 schools, one middle school, and one high school. All of the students in the study received the majority of their instruction in the regular education classroom.

Predictor and Criterion Variables

Predictor variables. The predictor variables in this study were two CBM measures in reading: reading aloud and maze selection. Reading aloud passages consisted of passages varying in length from 415 to 460 words (Grade 3), 394 to 424 words (Grade 5), 571 to 909 words (Grades 8 and 10), 1,277 to 1,481 words (Common Passage). An unnumbered copy of the passage was given to the student; a numbered copy to the administrator to allow for immediate scoring. Students read aloud for one minute, and the number of words read correctly was recorded.

Maze selection passages were created from the reading aloud passages using procedures outlined in L.S. Fuchs, Fuchs, Hamlett, and Ferguson (1992). The first sentence of every passage was left intact. Every seventh word thereafter was deleted from the text and replaced with three word choices. If the word to be replaced was a proper noun or an article, the next appropriate word was selected and replaced with three word choices. Each multiple choice was composed of the correct choice and two distracters. Distracters were selected to be not semantically correct or visually similar to the correct choice. The distracters were no more than one letter longer or shorter than the correct choice. Students read silently through the maze passage for three minutes

and circled the word choices for each multiple-choice item. At the end of 1-, 2-, and 3- minutes, students made a slash mark through the word they were reading to allow for investigation of differences related to sample duration.

Two different scoring procedures were compared for the fall maze selection passages (for a detailed report of this analysis, see RIPM Technical Report #10). First, we compared a 2- vs. 3-consecutive error rule for determining when to stop scoring. Typically, to control for guessing, scoring for maze selection is stopped after a student makes 3 consecutive errors. Comparison of a 2- vs. 3- consecutive error rule revealed no differences in reliability and validity of fall data; thus, the 3-consecutive rule was adopted because it is the rule that has been used in previous research. Second, we compared scoring the number of correct choices vs. the number of correct minus incorrect choices. Scoring correct minus incorrect occasionally has been used in previous research as an additional control for guessing (e.g., see Deno, Maruyama, Espin, & Cohen, 1989; Espin, Deno, Maruyama, &Cohen, 1989). Again, results of our analyses (see RIPM Technical Report #10) revealed no differences in reliability and validity between the two scoring procedures; thus, we adopted scoring correct only because it was more efficient than scoring correct minus incorrect. For both reading aloud and maze selection, we used the students' mean score across the three passages in all analyses.

Passages were of two difficulty: *grade level* and *common level*. The grade level passage was designed to be at a level approximate to that of the students' grade level. Grades level passages for 3rd graders were at a 3rd-grade level, 5th-graders at a 5th-grade level, and 8th- and 10th-graders at a 7th-grade level (8th- and 10th-grade students read the same passages). The common passage was read by all students at every grade level (thus was *common* to all students), and was approximately at a 4th-grade level. The 3rd and 5th grade-level passages and the common

4th grade-level passages were selected from Project PROACT at Vanderbilt University. The 8th and 10th grade-level passages were adapted from human-interest stories published in the local newspaper.

Criterion variables. Criterion variables included a standardized achievement test and two state-standards tests. The standardized achievement test was the *Northwest Achievement Level Test (NALT)/ Measures of Academic Progress (MAP)* (Northwest Evaluation Association, 2003). The NALT is the paper and pencil version of the test; the MAP is the computer version. One of our participating districts used the NALT, the other the MAP. The test was given to students in Grades 3 and 5 in both districts, and Grade 8 in one district. Because the tests are vertically scaled and make use of Rasch unIT (RIT) scores, the scores across the two versions of the tests and across grades and skill levels can be compared.

The NALT/MAP is a nationally-normed, standardized achievement test. The test is adaptive in nature; that is, it is matched to the individual student's skill level. The first time students take the NALT form of the test, they complete a paper-and-pencil locator test to identify their skill level. Subsequent testing is based on the results of the locator test or on previous NALT performance. The MAP form of the test is dynamically adjusted as the student take the test, so that the difficulty of each item presented to the student depends on the accuracy of the students' previous answers.

The NALT/MAP reading test consists of several subtest including word recognition, vocabulary, literal comprehension, inferential comprehension, and evaluative comprehension. Districts may choose to give all or part of the reading test. We included in our analysis only scores for the subtests common to both districts, which were the Literal and Interpretive Comprehension subtests. The test questions are multiple-choice in format. Test-retest reliability

of the NALT/MAP as reported in the technical manual (Northwest Evaluation Association, 2003) ranges from .80 to .92 in grades 3 to 10, and is .77 for grade 2. Marginal reliability or the expected correlation between scores on two hypothetical tests taken by the same student across test forms is .90 to .94 for grades 2 to 10. Concurrent validity with the Stanford Achievement Test and the Iowa Test of Basic Skills ranges from r = .77 to .87 in grades 2 through 9. Correlations with state standards reading tests in Colorado, Indiana, Washington, and Wyoming ranges from r = .75 to .86 in grades 3 through 10, with most correlations above .80. The average standard error of measurement for the NALT/MAP is reported to be 3 to 3.5 RIT scores.

The two state tests were the *Minnesota Basic Standards Test (MBST)* and the *Minnesota Comprehensive Assessment (MCA).* The MBST is a criterion-referenced, minimum competency test. At the time of the study students took the test in 8th grade, and were required to pass the test for graduation. If students did not pass the test the first time they took it, they could retake it. The reading subtest of the MBST is an untimed test containing four passages (one narrative and three expository), each with at least 500 words. Passages are drawn from newspaper articles and are selected to be on topics of interest to adolescents. The difficulty levels of the passages are determined using Degrees of Reading Power (DRP; Touchstone Applied Sciences and Associates, 2006) which is designed to assess the reading difficulty of texts based on length of words, length of sentences and common or frequently used words. DRP scores range from 1 to 100. The average DRP scores for the MBST passages range from 64 to 67, and cannot be below 62 or above 69. Approximately 65% of the test questions focus on literal comprehension and 35% on inferential comprehension. Each test has 40 multiple-choice questions. Scores are scaled from 368 to 750. A score of 600 is needed for passing (Minnesota Department of Education, 2005).

The *Minnesota Comprehensive Assessment (MCA)* is a criterion-referenced test based on the curricula used in Minnesota and is administered to all Minnesota students in grades 3, 5, 7, 10, and 11 (Minnesota Department of Education, n.d.; Retrieved June 2, 2005, from http://education.state.mn.us/content/087687.doc). At the time of our study, both the MBST and MCA were given. Currently only the MCA is given. The MCA is given as a part of NCLB, and is used to determine Adequate Yearly Progress (AYP) for schools. The purpose is to measure student progress toward achieving excellence in high standards, as well as to provide information about the range of achievement levels among all grades and levels of students. The MCA was developed on the basis of content expertise. Validity of the test is determined by the extent to which the content matches expert judgment.

The format of the MCA differs by grade level. The test consists of up to 8 passage selections. In Grades 3 and 5, informational, practical, and literary passages are used. Informational passages are nonfiction selections commonly found in periodicals, textbooks, etc. Practical selections are intended for specific applications such as recipes, how to instructions, advertisements, etc. Literary selections are fictional selections such as short stories, poems, excerpts from novels, etc. The passages for Grades 3 and 5 ranged from 250 to 1000 words. DRP levels for Grade 3 range from 40 to 56. A DRP level of 48 is considered to be a typical 3rd grade level passage. DRP levels for Grade 5 passages range from 44 to 64. A DRP of 54 is considered to be a typical 5th grade level passage. In Grade 10, passages are expository in nature and include contemporary essays, historical passages and technical selections. The passages are 400 to 1000 words in length, and have DRP levels ranging from 55 to 75. A DRP of 69 is considered to be typical for a 10th-grade level passage.

The questions on the MCA are 46 multiple-choice questions, each worth 1 point, and 3 constructed response questions, each worth 4 points. Questions are literal, interpretive, and evaluative in nature. The MCA uses scaled scores that range from 160 to 2260. Proficiency benchmarks are set for each grade level, with scores of 1420 or more being at or above grade level.

Procedure

Schools were recruited for participation based on their interest in progress monitoring student performance in reading. Consent forms were sent home with all students in grades 3rd, 5th, 8th, and 10th. Consent forms were sent home in English, Spanish, Hmong, and Somali depending upon the home language. Student assent was also obtained. All students with both parent consent and student assent were included in the study.

Data collection rounds for the curriculum-based measures took place in Fall, Winter, and Spring of 2004-2005. Each data collection round consisted of two sessions, one week apart. During the two weeks, students completed the same passage both as a reading aloud and maze selection passage; however, there was always a week between administration of the passages. The order in which the tasks were completed was counterbalanced. Thus, half of the students completed the passages in the following order (see Table 2): common maze selection, grade-level reading aloud (Week 1), grade-level maze selection, common reading aloud (Week 2). The other half completed the passages in the following order: grade-level maze selection, common reading aloud (Week 1), common maze selection, grade-level reading aloud (Week 2). In sum, a total of 6 passages were administered, one week as reading aloud or maze selection, and the next week in the other format. The order of passages was the same for all students; just the response format was changed. In other words, students read passages 1, 2, 3, 4, 5, and 6 in the same order,

but for half of the students, 1, 2 and 3 appeared as reading aloud the first week and maze selection the second week, and for the other half, maze selection the first week and reading aloud the second. Students completed the same passages in the same order across Fall, Winter, and Spring testing administrations.

Graduate students were trained to administer and score all Curriculum-Based Measures of reading prior to the start of data collection. Data collectors and scorers attended two separate two-hour training sessions. A subset of the data collectors who collected reading aloud data also scored the maze selection probes. During the training sessions, data collectors practiced administering and scoring reading aloud and maze selection passages. Data collectors had to reach 90% scoring accuracy on three reading aloud or maze selection passages to begin scoring. If 90% was not reached, procedures for scoring were reviewed and practice continued until the data collector met the 90% criterion.

During data collection, scoring accuracy for reading aloud was checked daily. Data collectors tape-recorded the administration of two reading aloud samples each day. These tape-recordings were scored by the trainer and inter-scorer agreement was calculated by dividing the smaller score by the larger score. For maze selection, passages for every 20th student were checked for accuracy. Accuracy checks were done in the Fall, Winter, and Spring. Inter-scorer agreement between the trainer and the scorer was calculated by dividing the smaller score by the larger score. The average and range of inter-scorer agreement for reading aloud and maze selection for fall, winter, and spring data collection is presented in Table 3. The criterion variables were administered and scored by the districts (NALT/MAP) or the state (MBST). The MBST was given in the Winter. The NALT/MAP was given in the Spring.

Analyses

Alternate-form reliability was calculated by examining correlations among the three forms of reading aloud (common and grade level) or maze selection (common and grade level) for each student in the Fall, Winter, and Spring testing sessions. Concurrent validity was examined by calculating correlations between the Winter CBM and MBST scores for students in grade 8, Spring CBM and NALT/MAP scores for grades 3, 5, and 8, and Spring CBM and MCA scores for grades 3, 5, and 10. Predictive validity was examined by calculating correlations between Fall CBM and Spring NALT/MAP and MCA scores for the respective grades. To examine whether relations differed for skill level, the linearity of the relation between the predictor and criterion variables was statistically tested at each grade level. A linear relation would imply that the relation between the predictor and criterion remained the same at various skill levels. A non-linear relation would imply a difference in the magnitude of the relation for different skill levels. Finally, the extent to which the measures reflected growth was examined in two ways. First, within each grade, average growth for students from Fall to Winter to Spring was calculated on each measure using latent-growth model analyses. Second, growth across grades on each measure was examined using scores across grades on the Fall, Winter, and Spring testing.

Results

Means and Standard Deviations

The means and standard deviations for reading aloud and maze selection measures in Fall, Winter, and Spring testing for grades 3, 5, 8, and 10 are reported in Table 4. The average NALT/MAP Literal Comprehension subtest RIT scale scores at each grade level were: 199.91 (*SD* = 13.82, *n* = 107) for Grade 3, 214.10 (*SD* = 16.05, *n* = 126) for Grade 5, and 224.44 (*SD* = 11.36, *n* = 43) for Grade 8. The average NALT/MAP Interpretive Comprehension subtest RIT

scale scores at each grade level were: 201.76 (*SD* = 14.09, *n* = 107) for Grade 3, 212.68 (*SD* = 15.39, *n* = 126) for Grade 5, and 225.79 (*SD* = 13.26, *n* = 43) for Grade 8. The average NALT/MAP combined subscale RIT scale scores at each grade level were: 401.66 (*SD* = 26.00, *n* = 107) for Grade 3, 426.79 (*SD* = 29.33, *n* = 126) for Grade 5, and 450.23 (*SD* = 21.17, *n* = 43) for Grade 8. The average overall reading RIT scale scores on the NALT/MAP (different subtests included for each district) ranged from 140 to 280. Average scores at each grade level were: 200.83 (*SD* = 12, *n* = 107) for Grade 3, 213.39 (*SD* = 14.66, *n* = 126) for Grade 5, and 225.12 (*SD* = 10.58, *n* = 43) for Grade 8. The average scale scores on the MCAs ranged from 360 to 2330. Average scores at each grade level were: 1523.37 (*SD* = 170.61, *n* = 104) for Grade 3, 1561.81 (*SD* = 284.53, *n* = 127) for Grade 5, and 1539.43 (*SD* = 17.24, *n* = 141) for Grade 10. The mean score on the MBST for 8th grade students was 636.10 (*SD* = 39.63, *n* = 86).

Alternate-form Reliability

Alternate-form reliabilities for reading aloud and maze selection across grade levels are reported in Table 5 and graphed in Figure 2. Reliability coefficients ranged from *r* = .74 to .94 in fall, *r* = .74 to .95 in winter, and *r* = .77 to .95 in spring. The reliability coefficients were above .80, with the exception of five reliability coefficients in fall, three reliability coefficients in winter, and one reliability coefficient in spring. Reliability coefficients tended to be larger for reading aloud (all above .90 with the exception of fall grade level reading aloud for Grade 8) than for maze selection (most in the .80s). Few differences were seen for alternate-form reliabilities across grade levels or between common and grade-level passages, with the exception that the coefficients for the grade-level maze selection passages were slightly lower for 8th-graders than for other grades. There were consistent effects related to the duration of maze selection. For both common and grade-level passages, and across all grade levels, 2- and 3-

minute maze selection produced higher alternate-form reliability coefficients than 1-minute maze selection, although even the 1-minute passages tended to have coefficients above .80. Virtually no differences were noted in alternate-form reliability between 2 and 3 minutes maze selection.

Concurrent Validity

Concurrent validity was examined by calculating correlations between the CBM measures and the NALT/MAP (grades 3, 5, and 8 only) and the MCAs (grades 3, 5, and 10 only). Correlations between the CBM measures and the NALT/MAP are reported in Table 6 and graphed in Figure 3. The pattern of results was similar across both reading aloud and maze selection, although correlations tended to be somewhat stronger for reading aloud than maze selection at each grade level; however, the magnitude of the differences was small. Perhaps the most noticeable outcome was the drop in correlations at the 8th-grade level. This drop occurred across reading aloud and maze selection, and for both grade-level and common passages. Whereas correlations for 3rd grade ranged from .58 to .67 and in 5th grade from .66 to .76, correlations in 8th grade ranged, with one exception, from .42 to .49. Only grade level reading aloud produced a correlation above .49 (r = .61). In general, the pattern of relations tended to be the same for both common and grade-level passages, with the exception of a stronger correlation seen for reading aloud in 8th grade. (However, as will be seen, this finding was not replicated with the MBST analyses.) With respect to duration of the maze selection measure, for both common and grade-level passages, correlations tended to be higher for 2- and 3-minute passages than for 1-minute passages. Correlations for 2 and 3 minutes were very similar.

Examination of the scattergrams of the relation between predictor and criterion variables reveals generally linear relations across most measures and most grades (see Figure 4).

Concurrent validity was also examined by calculating correlations between Spring CBM and MCA scores for 3, 5 and 10th grade students (see Table 7 and Figure 5). Correlations with the MCA tended to be similar to those with the NALT/MAP, and a similar pattern of results emerged with respect to measure, grade level, duration, difficulty level, and skill level. As with the NALT/MAP, the pattern of results was similar across both reading aloud and maze selection, with most correlations between .63 and .76. As with the MAP/NALT, there was a drop in correlations for the secondary-level students (in this analysis, 10th-grade students). Whereas most correlations for 3rd- and 5th-graders were above .63, no correlations for the 10th graders exceeded .63. However, the correlations for the 10th graders in the MCA analyses were noticeable stronger (r = .55 to .62) than for the 8th-graders in the NALT/MAP analysis (r = .42 to .49 for all but one correlation of r =.61). As with the NALT/MAP, the pattern of relations tended to be the same for both common and grade-level passages, with the exception that at grade 3, correlations for grade-level maze selection were somewhat lower than for common maze selection or for reading aloud, but the magnitude of the differences was small. With respect to duration of the maze selection measure, for both common and grade-level passages, correlations tended to be higher for 2- and 3-minute passages than for 1-minute passages, and correlations for 2 and 3 minutes were very similar. This was the same pattern of results seen for the NALT/MAP.

Examination of the linearity of the relation between predictor and criterion variables revealed a linear relation between the CBM measures and the MCAs. This was true for both measures (both common and grade level) at all durations and at all grade levels. This linear relationship is illustrated in Figure 6.

Finally, for the 8th-grade students only, the relationship between performance on the CBM measures and the MBST was examined. The predictive (fall CBM) and concurrent (winter

CBM) validity coefficients are shown in Table 8. The predictive validity coefficients ranged from *r* = .34 to .47 and the concurrent validity coefficients ranged from *r* = .41 to .49.

Predictive Validity

Predictive validity was examined by calculating correlations between performance on the CBM measures in the Fall and performance on the NALT/MAP and MCAs in the Spring. The predictive validity coefficients are presented in Tables 9 and 10. The predictive validity coefficients for reading aloud (common and grade level) and the NALT/MAP ranged from *r* = .48 to .73. In general, the correlation coefficients for reading aloud were slightly stronger than maze selection. The strongest coefficients were found for Grade 5 (.72 and .73 for common and grade level respectively), followed by Grade 3 (.68 and .66 for common and grade level respectively), and finally Grade 8 (.48 and .57 for common and grade level respectively). For maze selection, the validity coefficients ranged from *r* = .23 to .70, depending on grade level and difficulty level. The strongest set of coefficients was found at Grade 5 with most in the mid .60s to .70 for both common and grade level, followed by Grade 3 with coefficients ranging from *r* = .57 to .63, and lastly Grade 8 (*r* = .23 to .45). The correlation coefficients for grade level maze selection were unusually low for Grade 8 (.23 to .26).

Predictive validity was also examined using Fall CBM measures and the spring MCAs. As seen with the NALT/MAP, the correlation coefficients for reading aloud were slightly stronger for reading aloud in Grades 3 and 5. In Grade 10, common maze selection correlation coefficients were slightly stronger than reading aloud. The reading aloud coefficients ranged from *r* = .69 to .74 for students in Grades 3 and 5, and from *r* = .55 to .59 for students in Grade 10. For maze selection, the coefficients ranged from *r* = .61 to .66 for Grade 3, *r* = .63 to .72 for

Grade 5, and r = .57 to .62 for Grade 10. The grade level maze selection coefficients were slightly lower than the common maze selection coefficients for Grades 5 and 10.

Growth

Sensitivity to growth was examined in two ways: growth for students from Fall to Winter to Spring within grade, and growth across students at different grade levels. Estimated means based on an HLM growth analysis from Fall to Winter to Spring on the CBM scores are reported in Table 11. The estimated changes in scores from Fall to Winter to Spring are presented in Figure 7 for reading aloud and Figures 8 and 9 for maze selection common and grade level respectively. For the reading aloud measure, results for both the common and grade level passages revealed intercept effects (see Tables 12 and 13). On the common passages, 10th-graders started out reading approximately 105 more words per minute than 3rd- graders (t = -20.49, p < .001), 56 more words per minute than 5th- graders (t = -11.59, p < .001), and 11 more words per minute than 8th-graders (t = -1.99, p = .05). On the grade level passages, 10th-graders started out reading approximately 42 more words per minute than 3rd- graders (t = -9.26, p < .001) and 12 more words per minute than 8th-graders (t = -2.49, p = .01). No significant differences were found between 5th- and 10th-graders (t = -.88, p = .38). Analyses comparing intercepts for 3rd- and 5th-graders, 3rd- and 8th-graders, and 5th- and 8th-graders were not conducted.

As is illustrated in Figure 7, significant growth occurred at each grade level on both the common and grade-level passages although the amount of growth differed by grade level (see Tables 12 and 13). On the common passage, the greatest growth is seen for 3rd- and 5th-graders, each gaining on average 31 WRC over the course of the year, compared to 20 and 18 WRC for 8th- and 10th-graders respectively. Growth for the 10th-grade students differed significantly from

that of the 3rd- and 5th-grade students ($t = 6.31$, $p < .001$ and $t = 6.23$, $p < .001$, respectively). A similar pattern of results emerges on the grade-level passages, with 3rd- and 5th-graders gaining 28 and 29 WRC over the course of the year respectively, compared to 20 and 16 for the 8th and 10th-grade students respectively. Again, growth for the 3rd- and 5th-graders differed significantly from that of the 10th-graders ($t = 6.50$, $p < .001$ and $t = 6.95$, $p < .001$, respectively). No significant differences in growth were found between 8th- and 10th-graders on common passages ($t = 1.33$, $p = .19$) or grade level passages ($t = 1.90$, $p = .06$). Interestingly, the magnitude of the growth across the school year was similar on both the common and grade-level passages, and within 3 WRC on the two levels. Analyses comparing growth for 3rd- and 5th-graders, 3rd- and 8th-graders, and 5th- and 8th-graders were not conducted.

For the maze selection measure, results for both the common and grade level passages also revealed intercept effects (see Tables 14-19). Tenth graders started out with more correct maze selections than 3rd-, 5th-, and 8th-graders on both the common and grade level passages at 1, 2, and 3 minutes. Tenth graders made approximately 9 more maze selections than 3rd-graders ($t = -18.30$, $p < .001$), 6 more maze selections than 5th-graders ($t = -11.62$, $p < .001$), and 2 more maze selections than 8th-graders ($t = -3.20$, $p < .01$) on the common passage at 1 minute. The gap in performance remained fairly consistent across maze duration with 10th-graders making approximately 27 more maze selections than 3rd-graders ($t = -19.15$, $p < .001$), 17 more maze selections than 5th-graders ($t = -12.77$, $p < .001$), and 5 more maze selections than 8th-graders ($t = -3.33$, $p < .001$) on the common passage at 3 minute. The results for the grade level passage were very similar with the gap in performance being less between 10th-graders and the other grade levels. Analyses comparing intercepts for 3rd- and 5th-graders, 3rd- and 8th-graders, and 5th- and 8th-graders were not conducted.

Results of yearly growth for maze selection are presented in Figures 8 (common passages) and 9 (grade-level passages). In Figure 8, results for 1, 2, and 3 minutes of maze are presented. Perhaps most obvious, and not surprising, is that growth is easier to detect with longer duration samples on maze. Annual growth for a 1-minute sample ranged from 4.5 to 5.5 correct maze selections, which is relatively little growth over the course of several months. Fifth graders grew approximately .40 correct maze selections greater than 10th-graders ($t = 2.18$, $p = .03$). No significant differences in growth rates were found between 10th-graders and 8th-graders or 3rd-graders. At 3 minutes, growth was of greater magnitude, ranging from 11 correct maze selections to 13.5 correct maze selections across the year. Again, 5th-graders exhibited a steeper slope (1.24 correct maze selections greater) than 10th-graders ($t = 2.69$, $p < .01$). Analyses comparing growth for 3rd- and 5th-graders, 3rd- and 8th-graders, and 5th- and 8th-graders were not conducted.

Results for grade-level passages are similar to those for common passages (see Figure 9). As with the common passages, the magnitude of growth is larger at all grade levels for 3 minutes (ranging from 11 correct maze selections to 13.5 correct maze selections) than for 1 minute (ranging from 5 correct maze selections to 6 correct maze selections). Third graders demonstrated steeper growth (approximately .36 correct maze selections) than 10th-graders for 1 minute ($t = -2.00$, $p = .05$), while 5th-grade students tended to exhibit higher rates of growth (about 1.19 correct maze selection) than 10th-graders at 3 minutes ($t = 3.12$, $p < .01$). No significant differences were found between 10th-graders and 8th-graders or 5th-graders for 1 minute, or between 10th-graders and 8th-graders or 3rd-graders for 3 minutes. Again, analyses comparing growth for 3rd- and 5th-graders, 3rd- and 8th-graders, and 5th- and 8th-graders were not conducted. Interestingly, as with reading aloud, the magnitude of growth across time is similar for the common and grade-level passages.

Our second approach for examining sensitivity of the measures to growth was to examine differences in mean scores on the measures between grade levels. Figure 10 presents mean scores by grade for reading aloud (top figure) and maze selection (bottom 2 figures) common and grade-level passages. What clearly emerges from this figure is the sensitivity of the common passage to between-grade differences. For both reading aloud and maze selection, the common passage produces fairly linear growth rates from grade 3 to 10. For reading aloud, the growth curve is negatively accelerating in the upper grades; however, for maze selection, there is no leveling off of growth at the upper grades. What is more, for 3 minute maze selection, a steep and linear growth rate can be seen from grade 3 to 10.

Validity of Growth Rates

In our final analysis, we examined whether the fall to winter to spring growth rates produced by the CBM measures would be related to performance on external criteria. The assumption underlying this analysis was that students who were higher-performing readers, as measured by the NALT/MAP or the MCA, would exhibit steeper rates of growth over time than lower-performing readers. To examine this question, we selected maze selection 3 minutes (common and grade-level passages) and reading aloud, 1 minute (common and grade-level passages). Separate HLM analyses were run for each CBM reading measure, with NALT/MAP or MCA and grade level entered as Level 2 predictors. Results revealed that higher scores on the NALT/MAP and the MCA were associated with higher CBM intercept scores and with greater increases on the CBM measures over time for all measures except reading aloud, grade-level passage, which did not show a significant relation with the NALT at any grade level (see Tables 20-25 and Figures 11-17).

Discussion

In this study, we focused on the initial step in the development of a seamless and flexible system of CBM progress monitoring that would extend across grade levels from elementary to high-school levels. Specifically, we examined the reliability, validity, and sensitivity to growth of reading progress measures as indicators of general reading proficiency, and examined whether the reliability, validity and sensitivity to growth would be affected by type of measures, sample duration or difficulty level of the passage, and whether these factors would differ by grade level.

Alternate-form Reliability

With respect to alternate-form reliability, our results revealed that reliabilities were relatively good for all measures at all grade levels and all durations for both the common and grade-level passages. The majority of all reliabilities were above .80. As has been found in previous research at the secondary-school level (Espin et al., 2009; Ticha et al., 2009), reliabilities for reading aloud tended to be higher (all above r = .93) than for maze selection (most between r = .80 and .92), but reliabilities for maze selection were still within the range common for CBM measures. Also, as seen in previous research (Espin et al., 2009; Ticha et al., 2009), reliabilities for maze selection increased with time. In our study, differences were especially noticeable between 1 and 2 minutes, with smaller differences, if any, between 2 and 3 minutes. The consistency of our results across grade levels and across common and grade-level passages suggest that it may be necessary to use at least a 2-minute maze selection measure as an indicator of performance. As to the need for a 3-minute maze selection measure, our study differs from previous research, which was conducted at the middle-school level (Espin et al., 2009; Ticha et al., 2009), in which reliabilities also increased somewhat from 2 to 3 minutes; however, these increases were small. Finally, our results reveal few differences in the patterns of

results for common or grade-level passages, with the exception of somewhat lower alternate-form reliabilities for grade-level passages at 8th and 10th grade than for common passages.

In summary, our results suggest that both reading aloud and maze selection produce generally reliable scores across grade levels and type of passage, with reliabilities consistently above r =.80. Consistent with previous research, reading aloud produces consistently higher reliability coefficients than does maze selection. Also consistent with previous research, reliability for maze selection increases with duration. In our research, differences were consistently seen between 1 and 2 minutes, but not between 2 and 3 minutes. In previous research, differences were seen between 2 and 3 minutes. We would suggest that at least a 2-minute maze selection be used for progress monitoring. For other purposes, such as fall, winter, and spring testing or norming, districts might want to consider a 3-minute measure.

Validity

In terms of validity, perhaps the most obvious outcome was that the strength of the relations between the CBM and criterion measures dropped as student age increased in age. These results were consistent across type of measure, duration, and difficulty level. Given the results of previous cross age studies (Jenkins and Jewell, 1993; Yovanoff, Duesbery, Alonzo, and Tindal, 2005), we expected that a drop in validity coefficients might be seen with an increase in student age level, especially for reading aloud. However, the low magnitude of the correlations seen for the students in 8th grade is surprising, and somewhat hard to explain, especially given the fact that the 10th-grade students did not exhibit the same low correlations with the MCAs. In addition, recent research at the middle-school level (Espin et al., 2009; Ticha et al., 2009) has demonstrated much stronger validity coefficients, ranging from correlations of .75 to .89, with the same state standards test used in this study and with the broad reading cluster

of the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001). Although other middle-school studies have resulted in more modest correlations for middle-school students (e.g., Espin & Foegen, 1996; Yovanoff et al., 2005), even in these studies correlations typically ranged between .55 and .60. Our only explanation is that there was something unique to the particular sample of 8th graders participating in this study that contributed to the lower correlations. Supporting this hypothesis is the fact that at one of the fall group testing sessions, one student claimed loudly that research was just part of a government conspiracy, and that the students did not *have* to participate in the study. When students asked if they were required to participate, the researchers, or course, answered that they were, in fact, not required to participate, and could withdraw from the study at any time (adding, of course, that their participation would be greatly appreciated). Half of the students got up and left.

Within grade level, the relation between the CBM and criterion variables tended to remain similar along the continuum of CBM scores. In other words, the relation between the CBM scores and criterion variables tended to be linear in nature.

With respect to type of measure, there was no clear pattern of differences between reading aloud and maze selection, except for the fact that 1 minute reading aloud measures tended to result in larger validity coefficients than 1 minute maze selection at each grade level. With respect to difficult level of passage, common and grade-level passages produced similar results, a surprising finding for the students in 8th and 10th grade. We had assumed that students at these grade levels might hit a ceiling in their scores on a 4th grade passage, or that the passage would not be sensitive to differences in student performance at that level. Correlations revealed that the common passage functioned similar to the grade-level passage in terms of predicting student performance on various criterion variables.

Finally, with respect to duration, validity coefficients were found to increase with duration, with consistently stronger correlations found for 2- and 3-minute durations than for 1 minute. Small or no differences were seen between 2- and 3-minute durations, replicating earlier research (Espin et al., 2009; Ticha et al., 2009).

In sum, the validity of both reading aloud and maze selection as indicators of general performance was supported at the 3rd-, 5th-, and, to a lesser extent, 10th-grade levels. The correlations seen at 8th grade were too low to consider them supportive of the validity of the CBM measures at this grade level, although these correlations are not in accordance with previous research.

Growth rates

All measures reflected growth from fall to winter to spring. Surprisingly, the amount of change did not differ for the common and grade-level passages; that is, within grade, students gained similar amounts regardless of the difficulty level of the passage. Further, the amount of change seen was similar by grade level, with only small differences seen for reading aloud, and no differences for maze selection.

With sensitivity to grade level differences, results revealed that use of a common passage was better for indexing cross grade growth than a grade-level passage. The common passage produced fairly linear growth rates from grade 3 to 10. Upon reflection this outcome is not surprising, given that the difficulty level of the passage increases with grade for the grade-level passages. However, what is surprising is that the students in 8th and 10th grade did not “top out” on the 4th-grade passage. In terms of type of measure, the cross-grade growth rates for reading aloud tended to flatten out at the upper grades, replicating findings from previous research (MacMillan, 2000). Maze selection however, did not result in a negatively accelerating curve.

Growth rates for maze remained linear, and were appeared steeper for the 3-minute maze selection.

The validity of the growth rates produced by the measures was examined by exploring the relation between the growth rates from fall to winter to spring, and scores on the NALT/MAP and MCAs. Results generally revealed a significant relationship between growth on the reading aloud and maze selection measures at each grade level and performance on the criterion variables. The only exception to this finding was with reading aloud, grade level passages and the NALT/MAP where no relationship was found at any grade level. These results contrast somewhat with previous research at the secondary-school level (Espin et al., 2009; Ticha et al., 2009) in which reading aloud was shown to produce flat growth rates or growth rates that were not related to external criteria, while maze selection produced substantial growth rates that were related to external criteria. The present study differed from the earlier studies in the sense that growth in this study was determined by change in performance on the same set of passages across three times in a year. In the previous research, growth was based on weekly measurement on alternate forms of probes. There is a need for additional research on the sensitivity and validity of growth rates produced by CBM measures, especially for reading aloud, and especially at secondary-school level. Moreover, it is important to explore differences in growth rates produced by different data collection schedules (e.g., three times a year vs. weekly) and by parallel passages vs. use of the same passages.

Conclusion

In conclusion, the results of this study provide support for the reliability, validity, and sensitivity to growth of CBM reading aloud and maze selection measures at grades 3 and 5, and to a lesser extent, grade 10. Results of the present study did not support the validity of the

measures for 8th-grade students, but reliability and sensitivity to growth on the measures was good for these students. In our study, reading aloud and maze selection functioned similarly as indicators of performance and progress from grads 3 to 10. For both measures, the strength of the validity coefficients drops somewhat as students get older. Within grade, no or small skill level differences were seen on the measures, and those that were seen were only seen at the upper levels of the performance continuum. If teachers decide to use maze selection, a 2 or 3-minute measures seems to be more reliable and valid than a 1-minute measure at every grade level, although even 1-minute maze selection produced relatively reasonable reliability and validity coefficients, and reflected growth.

With respect to the idea of a seamless and flexible system of progress monitoring, our results clearly show that the best measure for reflecting growth across grade levels from elementary to high-school is a maze-selection, common passage, timed for 3 minutes. Reading aloud showed a slight deceleration in growth rates across grade levels, whereas maze selection resulted in linear growth rates. Only the common passage reflected consistent, linear growth across grade levels, and 3 minutes produced steeper growth rates than 1 or 2 minutes. Schools or districts could probe their students at the beginning of each school year with two 3-minute maze selection probes at a 4th-grade level, and use the results to evaluate growth over time for students in that school or district. If normative growth rates are established, teachers could examine these growth rates to determine to what extent students are growing across years relative to average- or similar-performing peers. If a student's growth across years is far below his or her peers, it may reflect a need to change the overall structure of that student's program to improve performance.

In conclusion, the results of our research support the notion that a seamless and flexible system of progress monitoring could be built to follow growth of students across years, and that,

at least from grades 3 to 10, a single measure can be used to index that growth.

Footnotes

[1] Percentages include preK-12, which increases the special education enrollment by 1%.

References

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

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

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62,* 497-514.

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

Espin, C.A., Wallace, T., Lembke, E., Campbell, H., & Long, J. (2009). Creating a progress measurement system in reading for secondary-school students: Monitoring progress towards meeting high stakes assessment standards. *Paper submitted for publication.*

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. *Exceptional Children,* 58, 436-450.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal comprehension measures. *Remedial and Special Education, 9,* 20-28.

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

MacMillan, P. (2000). Simultaneous measurement of reading growth, gender, and relative-age effects: Many-faceted Rasch applied to CBM reading scores. *Journal of Applied Measurement, 1,* 393-408.

Minnesota Department of Education. (2005). *Minnesota Basic Skills Tests technical manual*. Data Recognition Corporation.

Technical manual for the NWEA measures of academic progress and achievement levels tests. (2003). Portland, OR; Northwest Evaluation Association.

Tichá, R. Espin, C.A., & Wayman, M.M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading aloud and maze selection measures. *Learning Disabilities Research and Practice*, *24,* 132-142.

Touchstone Applied Science and Associates (2006). Degrees of reading power. Brewster, NY: Author.

Wayman, M.M., Wallace, T., Ives Wiley, H., Tichá, R., & Espin, C.A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41*, 85-120.

Wallace, T., Espin, C.A., McMaster, K., Deno, S.L, and Foegen, A. (2007). CBM progress monitoring within a standards-based system: Introduction to the special series. *Journal of Special Education, 41,* 66-67.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4-12.

Appendix

Reading Aloud Administration Directions
Maze Administration Directions

Reading Aloud Administration Directions

CBM READING
ADMINISTRATION DIRECTIONS

MATERIALS

1. Unnumbered copy of passage (student copy)
2. Numbered copy of passage (examiner copy—in student test packet)
3. Stopwatch
4. Clipboard
5. Red pen
6. Calculator

DIRECTIONS

1. Place the unnumbered copy in front of the student. Say "I'm going to have you read a story to me."

2. Place the numbered copy in front of you but shielded so the student cannot see what you record.

3. Say these specific directions to the student for the first passage.

   "When I say 'begin', start reading aloud at the top of this page."

   Point to the first word of the story, not to the title.

   "Read across the page."

   Demonstrate by pointing.

   "Try to read each word. If you come to a word you don't know, I'll tell it to you. If you get to the end of the passage before I say stop, start at the beginning of the passage again. Be sure to do your best reading. The title of the passage you will be reading is: ***(READ THE TITLE OF THE STORY OUT LOUD).*** Are there any questions?" ***(Pause)***

4. Point to the first word and say " **Ready? Begin**" Start your stopwatch when the student says the first word. If the student fails to say the first word of the passage after 3 seconds, tell them the word and mark it as incorrect, then start your stopwatch.

5. Follow along on your copy. Put a slash (/) through words read incorrectly (see scoring procedures).

6. If a student stops or struggles with a word for 3 seconds, tell the student the word and mark it as incorrect.

7. At the end of 1 minute, place a bracket ( ]) after the word that the student has just read, and say, **"Stop."**

8. Collect the first passage from the student; place the unnumbered copy of the second passage in front of the student. Administer this second passage with the following directions.

9. Say, "Now you're going to do the same thing with another story. Remember to do your best reading. The title of this story is ____. Any questions?"

COLLECT ALL MATERIALS

NOTE:

- Make sure to score the students' passages immediately after administering the measures.
- If you make an examiner's mistake during the administration of the reading aloud passage, tell the student to stop, restart your stopwatch, and have the student read from where they left off (or at the beginning of the next paragraph or sentence).

Adapted from CBM Administration and Scoring Module.
Shinn, M.R. (1989). Curriculum-based measurement: Assessing special children. New York: Guilford Press.

Maze Administration Directions

CBM MAZE
ADMINISTRATION DIRECTIONS

MATERIALS

1. Maze packet for each student.
2. Stopwatch.

DIRECTIONS

Say to the students: "Please write your first and last name, your teacher's name, the name of your school, the hour/period you have this class, and today's date on the top of your packet."

"Today we want you to read 3 short stories. The stories you are going to read have some places where you need to choose the correct word. You will read the story, and whenever you come to three words that are underlined and in dark print, you will circle the word that belongs in the sentence."

"Before you begin, we will do some examples. Look at the first page in your booklet. The first sentence says:

The snow was falling and the air was crisp. He put on his **trees / boots / houses** and walked to school."

"Circle the word that belongs in the sentence."

After 10 seconds: "The word boots belongs in the sentence, **He put on his boots**…. Circle the word boots."

Monitor the students for compliance.

Say to the students: "Now let's try sentence number two. The sentence says:

He was late, so he **map / see / ran** to catch the bus."

"Circle the word that belongs in the sentence."

After 10 seconds: "The word ran belongs in the sentence, **he ran to catch the bus.** Circle the word ran."

Monitor the students for compliance. Point to the word if necessary.

PASSAGE 1

Say to the students: "Please put your pencils down and listen to my directions." (Pause and monitor students for compliance.)

"Now you are going to do the same thing by yourself. You will read a story. Whenever you come to three words that are underlined and in dark print, circle the word that belongs in the sentence."

"Circle a word even if you're not sure of the answer. I cannot tell you any words, so do your best. If you make a mistake, don't erase, but put an X on the answer that you didn't want, circle the answer that you wanted, and move on."

Demonstrate for students. Put the words "He put on his trees/ boots / houses" on the board or overhead. Circle trees, X it out, and then circle boots.

"At the end of 1, 2, and 3 minutes, I will ask you to put a slash through the word that you are currently reading."

Demonstrate by putting a slash through the word put, that you wrote on the board or overhead.

"After you put a slash through the word you are on, you should continue reading. Continue working until I tell you to stop, or you reach the blank page. If you finish early, check your answers. You may begin when I tell you to. Are there any questions? Turn to page 2 in your booklet."

Monitor students to make sure they are on the first maze passage in their booklet.

Say to the students: "Remember to do the best you can. Pick up your pencils. Ready? Begin."

After 30 seconds, say: "Remember, circle a word, even if you are not sure of the answer."

At 1 minute, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 2 minutes, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 3 minutes, say: "Stop. Put a slash through the word you're reading and put your pencils down. Turn to the stop page."

PASSAGE 2

Say to the students: "Now you will do the same thing on another passage. Remember to circle the word that belongs in the sentence. Circle a word even if you are not sure of the answer. You may begin when I tell you to. Turn to the next page in your booklet."

Monitor students to make sure they are on the second maze passage in their booklet.

Say to the students: "Remember to do the best you can. Pick up your pencils. Ready? Begin."

After 30 seconds, say: "Remember, circle a word, even if you are not sure of the answer."

At 1 minute, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 2 minutes, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 3 minutes, say: "Stop. Put a slash through the word you're reading and put your pencils down. Turn to the stop page."

PASSAGE 3

Say to the students: "Now you will do the same thing on another passage. Remember to circle the word that belongs in the sentence. Circle a word even if you are not sure of the answer. You may begin when I tell you to. Turn to the next page in your booklet."

Monitor students to make sure they are on the third maze passage in their booklet.

Say to the students: "Remember to do the best you can. Pick up your pencils. Ready? Begin."

After 30 seconds, say: "Remember, circle a word, even if you are not sure of the answer."

At 1 minute, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 2 minutes, say: "Stop. Put a slash through the word that you are reading. Then continue."

At 3 minutes, say: "Stop. Put a slash through the word you're reading and put your pencils down. Turn to the stop page."

COLLECT ALL MATERIALS

NOTE:

- Don't forget to give the following prompt after 30 seconds has passed, "Remember, circle a word, even if you are not sure of the answer."
- If students ask you to identify a word, remind them to just do the best they can.
- It is very important for our reliability that you keep a close watch on the timing of the test. Have a back-up (clock or watch) available in case your stop watch doesn't work.

Table 1

Demographic Characteristics of Sample

	Urban District		Rural District	
	n	%	*n*	%
Males	131	48	108	46
Home Language				
English	193	70	234	100
Hmong	15	6	1	< 1
Spanish	50	18		
Somali	11	4		
Laotian	1	< 1		
Cambodian	1	< 1		
Ethiopian	1	< 1		
English Dialect	1	< 1		
Total	274		235	

Table 2

Order of Passage Administration

	Form A	Form B
Week 1	Common Maze Selection Grade Level Reading Aloud	Grade Level Maze Selection Common Reading Aloud
Week 2	Grade Level Maze Selection Common Reading Aloud	Common Maze Selection Grade Level Reading Aloud

Table 3
Interscorer Agreement for Fall, Winter, and Spring Reading Aloud and Maze Selection

Variable	*Average*	*Range*	*n*
Reading Aloud			
Fall	98%	87%-100%	13
Winter	98%	91%-100%	13
Spring	99%	92%-100%	10
Maze Selection			
Fall	91%	72%-100%	6
Winter	91%	69%-100%	5
Spring	99%	89%-100%	4

Table 4

Means and Standard Deviations for Reading Aloud and Maze Selection in Fall, Winter, and Spring Across Grade Levels

		Grade											
		3			5			8			10		
Measure	Time	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Reading aloud common	1	90.79	36.10	111	139.88	43.97	130	185.22	41.53	90	195.72	44.80	176
	2	108.60	39.34	107	156.18	47.47	129	195.16	41.91	86	207.94	44.11	164
	3	122.64	41.39	106	171.79	49.68	126	205.43	42.65	82	213.76	45.28	164
Reading aloud grade	1	103.52	36.40	111	142.53	41.77	130	134.66	31.70	90	146.54	36.91	178
	2	120.29	37.86	107	158.60	45.23	127	143.23	33.00	85	155.49	38.29	167
	3	132.76	37.70	106	172.22	46.56	125	153.70	35.93	82	163.06	38.10	164
Maze common 1 min	1	5.89	3.07	111	9.54	3.85	130	13.46	4.09	90	15.18	5.13	178
	2	8.48	3.51	107	13.01	4.56	127	16.62	4.33	85	18.49	5.65	168
	3	10.41	4.09	106	15.20	5.51	125	18.59	5.39	82	20.16	6.39	164
Maze common 2 min	1	11.39	5.92	111	18.19	7.15	130	26.20	7.55	90	29.83	9.91	178
	2	16.25	6.64	107	24.62	8.62	127	32.03	8.22	85	35.62	10.65	168
	3	19.46	7.32	106	28.40	10.32	125	34.91	10.34	82	38.54	11.90	164
Maze common 3 min	1	16.86	8.88	111	26.91	10.63	130	39.15	10.83	90	44.30	14.51	178
	2	23.77	9.70	107	35.46	12.66	128	46.64	11.97	85	52.20	15.49	168
	3	28.11	10.62	106	40.55	14.88	125	50.24	14.97	82	55.75	17.61	164
Maze grade 1 min	1	6.99	3.28	111	10.28	4.13	129	10.46	3.59	90	11.82	4.12	178
	2	9.90	3.74	107	14.15	4.66	129	13.15	3.99	86	15.28	5.18	167
	3	11.77	4.19	106	16.31	5.47	126	15.99	4.52	82	17.45	5.75	164
Maze grade 2 min	1	13.38	5.87	111	19.38	7.58	129	20.04	6.78	90	23.04	8.61	178
	2	18.64	6.39	107	25.90	8.26	129	25.39	7.65	86	28.89	9.84	167
	3	21.62	7.42	106	29.76	10.09	126	29.65	8.62	82	32.31	10.25	164
Maze grade 3 min	1	19.41	8.27	111	27.89	10.84	130	29.91	9.89	90	34.12	12.37	178
	2	26.42	9.12	107	36.49	11.92	129	36.60	10.68	86	41.35	13.61	167
	3	30.17	10.76	106	41.54	14.45	126	41.68	11.69	82	45.52	14.06	164

Table 5
Alternate-form Reliability for Reading Aloud and Maze Selection Across Time and Grade Levels

Reliability CBM Fall, Winter, and Spring

		Reading Aloud		Maze					
		Common	Grade	Common			Grade		
		Correct		Correct			Correct		
		1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Fall	Grade								
	3	.94**	.91**	.78**	.85**	.87**	.76**	.81**	.85**
	5	.93**	.93**	.77**	.85**	.88**	.82**	.86**	.88**
	8	.91**	.89**	.82**	.87**	.88**	.79**	.85**	.85*
	10	.93**	.92**	.85**	.90**	.91**	.74**	.84**	.83**
Winter									
	3	.94**	.93**	.76**	.86**	.88**	.78**	.86**	.90**
	5	.94**	.93**	.84**	.90**	.91**	.87**	.90**	.93**
	8	.95**	.92**	.82**	.89**	.91**	.74**	.81**	.83**
	10	.94**	.92**	.86**	.92**	.93**	.86**	.90**	.89**
Spring									
	3	.93**	.93**	.80**	.86**	.88**	.80**	.85**	.85**
	5	.95**	.95**	.85**	.90**	.92**	.87**	.91**	.93**
	8	.94**	.93**	.87**	.92**	.94**	.77**	.83**	.84**
	10	.93**	.95**	.85**	.93**	.95**	.82**	.87**	.87**

p** < .01

Table 6
Correlations between Reading Aloud and Maze Selection and the NALT/MAP across Grade Levels

Concurrent Validity, Spring 2005 CBM with Spring 2005 NALT/MAP
Comprehension subscales combined

	Reading Aloud		Maze					
	Common Mean	Grade Mean	Common Mean			Grade Mean		
	Correct		Correct			Correct		
	1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Grade								
3	.67	.66	.58	.63	.64	.58	.63	.64
5	.76	.72	.66	.71	.72	.67	.71	.74
8	.49	.61	.45	.44	.44	.42	.44	.44
10								

Table 7
Correlations between Reading Aloud and Maze Selection and the MCAs across Grade Levels

Concurrent Validity, **Spring 2005 CBM** with **Spring 2005 MCA**

	Reading Aloud		Maze					
	Common Mean	Grade Mean	Common Mean			Grade Mean		
	Correct		Correct			Correct		
	1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Grade								
3	.71**	.69**	.63**	.68**	.70**	.58**	.64**	.67**
5	.72**	.67**	.69**	.73**	.76**	.69**	.72**	.76**
8								
10	.58**	.62**	.55**	.58**	.58*	.56**	.60**	.62**

p** < .01
Note. Correlations for 8th grade are not available.

Table 8
Correlations between Reading Aloud and Maze Selection and the MBST Reading for Grade 8

Predictive and Concurrent Validity, **Fall 2004 and Winter 2005 CBM** with **Winter 2005 MBST Reading (*n* = 85)**

	Reading Aloud		Maze					
	Common Mean	Grade Mean	Common Mean			Grade Mean		
	Correct		Correct			Correct		
	1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Grade 8								
Fall	.37**	.34**	.46**	.47**	.46**	.38**	.39**	.41**
Winter	.45**	.43**	.48**	.46**	.49**	.41**	.42**	.44**

p** < .01

Table 9
Correlations between Fall Reading Aloud and Maze Selection and the NALT/MAP across Grade Levels

Predictive validity, Fall 2004 CBM with Spring 2005 NALT/MAP
Comprehension subscales combined

	Reading Aloud		Maze					
	Common Mean	Grade Mean	Common Mean			Grade Mean		
	Correct		Correct			Correct		
	1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Grade								
3	.68**	.66**	.57**	.62**	.63**	.59**	.59**	.60**
5	.72**	.73**	.60**	.64**	.65**	.66**	.67**	.70**
8	.48**	.57**	.40**	.45**	.40**	.26	.23	.23
10								

p** < .01
Note. Correlations for 10th grade are not available.

Table 10
Correlations between Fall Reading Aloud and Maze Selection and the MCAs across Grade Levels

Predictive Validity, **Fall 2004 CBM** with **Spring 2005 MCA**

	Reading Aloud		Maze					
	Common Mean	Grade Mean	Common Mean			Grade Mean		
	Correct		Correct			Correct		
	1 minute		1 min	2 min	3 min	1 min	2 min	3 min
Grade								
3	.74**	.71**	.63**	.65**	.66**	.61**	.63**	.65**
5	.71**	.69**	.63**	.66**	.68**	.66**	.68**	.72**
8								
10	.55**	.59**	.61**	.62**	.61**	.57**	.58**	.60**

p** < .01
Note. Correlations for 8th grade are not available.

Table 11
Reading Aloud and Maze Selection Mean HLM Estimates from Fall to Winter to Spring

		Grade			
		3	*5*	*8*	*10*
Measure	Time	*M*	*M*	*M*	*M*
Reading aloud common	1	91.83	140.11	185.57	196.39
	2	107.45	155.31	196.00	205.29
	3	123.06	170.51	206.42	214.2
Reading aloud grade	1	104.78	142.84	134.63	146.6
	2	119.13	157.29	144.61	154.56
	3	133.48	171.75	154.58	162.52
Maze common 1 min	1	6.04	9.76	13.67	15.42
	2	8.26	12.54	16.27	17.80
	3	10.48	15.33	18.86	20.19
Maze common 2 min	1	11.76	18.63	26.70	30.24
	2	15.71	23.67	31.14	34.44
	3	19.67	28.72	35.58	38.64
Maze common 3 min	1	17.43	27.49	39.82	44.91
	2	22.95	34.22	45.46	50.41
	3	28.48	40.96	51.09	55.91
Maze grade 1 min	1	7.19	10.56	10.46	12.00
	2	9.57	13.52	13.27	14.75
	3	11.95	16.48	16.08	17.49
Maze grade 2 min	1	13.82	19.83	20.26	23.39
	2	17.92	24.91	25.1	27.85
	3	22.03	29.98	29.95	32.3
Maze grade 3 min	1	20.04	28.51	30.22	34.55
	2	25.39	35.17	36.17	40.03
	3	30.74	41.83	42.12	45.5

Table 12
HLM Estimates for Common Reading Aloud

Reading Aloud Common Correct

Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		196.39	3.1555	505	62.24	<.0001
linear		8.9039	0.6679	477	13.33	<.0001
grade	3	-104.56	5.1032	504	-20.49	<.0001
grade	5	-56.2807	4.8545	504	-11.59	<.0001
grade	8	-10.82	5.4427	504	-1.99	0.0474
grade	10	0	.	.	.	.
linear*grade	3	6.7124	1.0639	474	6.31	<.0001
linear*grade	5	6.2949	1.01	476	6.23	<.0001
linear*grade	8	1.5232	1.1494	479	1.33	0.1857
linear*grade	10	0	.	.	.	.

Reading Aloud Common Correct

Grade	Intercept	Slope
3	91.83	15.62
5	140.11	15.20
8	185.57	10.43
10	196.39	8.90
Mean	153.47	12.54

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	91.83	140.1093	185.57	196.39
2	107.4463	155.3081	195.9971	205.2939
3	123.0626	170.5069	206.4242	214.1978

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	6353.83	<.0001
linear	1	477	974.29	<.0001
grade	3	504	160.76	<.0001
linear*grade	3	476	20.18	<.0001

Table 13
HLM Estimates for Grade Level Reading Aloud

Reading Aloud Grade Correct Mean

Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		146.6	2.7905	504	52.54	<.0001
linear		7.9613	0.614	477	12.97	<.0001
grade	3	-41.8229	4.5147	504	-9.26	<.0001
grade	5	-3.7601	4.2947	504	-0.88	0.3817
grade	8	-11.9706	4.8153	504	-2.49	0.0132
grade	10	0	.	.	.	.
linear*grade	3	6.3915	0.9828	474	6.5	<.0001
linear*grade	5	6.4917	0.9347	476	6.95	<.0001
linear*grade	8	2.0164	1.0619	479	1.9	0.0582
linear*grade	10	0	.	.	.	.

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	6022.18	<.0001
linear	1	476	987.28	<.0001
grade	3	504	31.82	<.0001
linear*grade	3	476	22.54	<.0001

Reading Aloud Grade Correct

Grade	Intercept	Slope
3	104.78	14.35
5	142.84	14.45
8	134.63	9.98
10	146.60	7.96
Mean	132.21	11.69

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	104.7771	142.8399	134.6294	146.6
2	119.1299	157.2929	144.6071	154.5613
3	133.4827	171.7459	154.5848	162.5226

Table 14
HLM Estimates for Common Maze Selection (1 min)

Maze	Common	Mean Correct Error 3	1 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		15.4168	0.3166	504	48.69	<.0001
linear		2.3862	0.1203	488	19.84	<.0001
grade	3	-9.372	0.5122	504	-18.3	<.0001
grade	5	-5.6612	0.4872	504	-11.62	<.0001
grade	8	-1.7466	0.5464	504	-3.2	0.0015
grade	10	0	.	.	.	.
linear*grade	3	-0.1674	0.1926	484	-0.87	0.3852
linear*grade	5	0.3991	0.1831	487	2.18	0.0298
linear*grade	8	0.2104	0.208	490	1.01	0.3122
linear*grade	10	0	.	.	.	.

Maze Common Mean Correct Error 3, 1 min

Grade	Intercept	Slope
3	6.04	2.22
5	9.76	2.79
8	13.67	2.60
10	15.42	2.39
Mean	11.22	2.50

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	6.0448	9.7556	13.6702	15.4168
2	8.2636	12.5409	16.2668	17.803
3	10.4824	15.3262	18.8634	20.1892

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3370.71	<.0001
linear	1	487	1174	<.0001
grade	3	504	127.37	<.0001
linear*grade	3	487	2.99	0.0308

Table 15
HLM Estimates for Common Maze Selection (2 min)

Maze	Common	Mean Correct Error 3	2 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		30.235	0.6051	504	49.97	<.0001
linear		4.201	0.2117	487	19.85	<.0001
grade	3	-18.479	0.9788	504	-18.88	<.0001
grade	5	-11.61	0.9311	504	-12.47	<.0001
grade	8	-3.5368	1.0441	505	-3.39	0.0008
grade	10	0	.	.	.	.
linear*grade	3	-0.2429	0.3389	483	-0.72	0.4739
linear*grade	5	0.8451	0.3223	485	2.62	0.009
linear*grade	8	0.2377	0.366	489	0.65	0.5163
linear*grade	10	0	.	.	.	.

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3492.17	<.0001
linear	1	486	1183.25	<.0001
grade	3	504	137.53	<.0001
linear*grade	3	486	3.58	0.0138

Maze Common Mean Correct Error 3, 2 min

Grade	Intercept	Slope
3	11.76	3.96
5	18.63	5.05
8	26.70	4.44
10	30.24	4.20
Mean	21.83	4.41

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	11.756	18.625	26.6982	30.235
2	15.7141	23.6711	31.1369	34.436
3	19.6722	28.7172	35.5756	38.637

Table 16
HLM Estimates for Common Maze Selection (3 min)

Maze	Common	Mean Correct Error 3	3 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		44.9148	0.8872	505	50.62	<.0001
linear		5.4953	0.3036	488	18.1	<.0001
grade	3	-27.4886	1.4353	504	-19.15	<.0001
grade	5	-17.4292	1.3653	504	-12.77	<.0001
grade	8	-5.0962	1.5311	505	-3.33	0.0009
grade	10	0	.	.	.	.
linear*grade	3	0.03196	0.4861	484	0.07	0.9476
linear*grade	5	1.2429	0.4623	486	2.69	0.0074
linear*grade	8	0.1415	0.525	490	0.27	0.7876
linear*grade	10	0	.	.	.	.

Maze Common Mean Correct Error 3, 3 min		
Grade	Intercept	Slope
3	17.43	5.53
5	27.49	6.74
8	39.82	5.64
10	44.91	5.50
Mean	32.41	5.85

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	17.4262	27.4856	39.8186	44.9148
2	22.95346	34.2238	45.4554	50.4101
3	28.48072	40.962	51.0922	55.9054

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3580.72	<.0001
linear	1	486	1011.41	<.0001
grade	3	504	142.57	<.0001
linear*grade	3	486	2.94	0.0327

Table 17
HLM Estimates for Grade Level Maze Selection (1 min)

Maze	Grade	Mean Correct Error 3	1 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		12.0038	0.2914	505	41.2	<.0001
linear		2.7422	0.1127	485	24.34	<.0001
grade	3	-4.8173	0.4713	504	-10.22	<.0001
grade	5	-1.4417	0.4485	504	-3.21	0.0014
grade	8	-1.5454	0.5027	504	-3.07	0.0022
grade	10	0	.	.	.	.
linear*grade	3	-0.3613	0.1803	482	-2	0.0456
linear*grade	5	0.2168	0.1713	485	1.27	0.2061
linear*grade	8	0.07102	0.1947	487	0.36	0.7155
linear*grade	10	0	.	.	.	.

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3194.19	<.0001
linear	1	485	1596.5	<.0001
grade	3	504	35.19	<.0001
linear*grade	3	484	3.2	0.0232

Maze Grade Mean Correct Error 3, 1 min

Grade	Intercept	Slope
Grade	Intercept	Slope
3	7.19	2.38
5	10.56	2.96
8	10.46	2.81
10	12.00	2.74
Mean	10.05	2.72

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	7.1865	10.5621	10.4584	12.0038
2	9.5674	13.5211	13.27162	14.746
3	11.9483	16.4801	16.08484	17.4882

Table 18
HLM Estimates for Grade Level Maze Selection (2 min)

Maze	Grade	Mean Correct Error 3	2 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		23.3892	0.5622	504	41.6	<.0001
linear		4.456	0.1847	486	24.12	<.0001
grade	3	-9.5675	0.9094	504	-10.52	<.0001
grade	5	-3.5557	0.8653	504	-4.11	<.0001
grade	8	-3.1342	0.97	504	-3.23	0.0013
grade	10	0	.	.	.	.
linear*grade	3	-0.3536	0.2956	483	-1.2	0.2322
linear*grade	5	0.6175	0.2808	486	2.2	0.0283
linear*grade	8	0.3927	0.3193	488	1.23	0.2193
linear*grade	10	0	.	.	.	.

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3170.55	<.0001
linear	1	486	1708.73	<.0001
grade	3	504	37.01	<.0001
linear*grade	3	486	3.73	0.0114

Maze Grade Mean Correct Error 3, 2 min

Grade	Intercept	Slope
Grade	Intercept	Slope
3	13.82	4.10
5	19.83	5.07
8	20.26	4.85
10	23.39	4.46
Mean	19.32	4.62

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	13.8217	19.8335	20.255	23.3892
2	17.9241	24.907	25.1037	27.8452
3	22.0265	29.9805	29.9524	32.3012

Table 19
HLM Estimates for Grade Level Maze Selection (3 min)

Maze	Grade	Mean Correct Error 3	3 min			
Effect	Grade	Estimate	SE	DF	T Value	P value
Intercept		34.5499	0.8061	504	42.86	<.0001
linear		5.4765	0.2502	487	21.89	<.0001
grade	3	-14.5063	1.304	504	-11.12	<.0001
grade	5	-6.042	1.2404	504	-4.87	<.0001
grade	8	-4.3267	1.3909	504	-3.11	0.002
grade	10	0	.	.	.	.
linear*grade	3	-0.1303	0.4003	484	-0.33	0.745
linear*grade	5	1.1859	0.3799	486	3.12	0.0019
linear*grade	8	0.4729	0.4324	489	1.09	0.2746
linear*grade	10	0	.	.	.	.

Effect	NUM DF	DEN DF	F Value	P Value
Intercept	1	504	3314.86	<.0001
linear	1	486	1499.2	<.0001
grade	3	504	41.72	<.0001
linear*grade	3	486	4.33	0.005

Maze Grade Mean Correct Error 3, 3 min

Grade	Intercept	Slope
Grade	Intercept	Slope
3	20.04	5.35
5	28.51	6.66
8	30.22	5.95
10	34.55	5.48
Mean	28.33	5.86

Time	Grade 3	Grade 5	Grade 8	Grade 10
1	20.0436	28.5079	30.2232	34.5499
2	25.3898	35.1703	36.1726	40.0264
3	30.736	41.8327	42.122	45.5029

Table 20
HLM Estimates for NALT/MAP Predicting Common Reading Aloud

READING COMMON <- NALT Subscale Total

Effect	Estimate	SE	df	t	p
Average Intercept	-420.63	26.4473	274	-15.9	<.0001
Average Linear Slope	-1.4315	6.9429	272	-0.21	0.8368
Total_Subscale*Intercept	1.3098	0.06269	274	20.89	<.0001
Total_Subscale*Linear	0.03844	0.01645	272	2.34	0.0202

Table 21
HLM Estimates for NALT/MAP Predicting Common Maze (3 min)

MAZE COMMON <- NALT Subscale Total

Effect	Estimate	SE	df	t	p
Average Intercept	-99.1112	7.0399	274	-14.08	<.0001
Average Linear Slope	-9.2444	3.1328	273	-2.95	0.0034
Total_Subscale*Intercept	0.2974	0.01669	274	17.82	<.0001
Total_Subscale*Linear	0.0364	0.007424	273	4.9	<.0001

Table 22
HLM Estimates for NALT/MAP Predicting Grade Level Maze (3 min)

MAZE GRADE <- NALT Subscale Total

Effect	Grade	Estimate	SE	df	t	p
Intercept		-63.7898	7.4761	273	-8.53	<.0001
linear		-12.4182	2.2423	274	-5.54	<.0001
GRADE	3	-0.9332	1.5425	272	-0.6	0.5457
GRADE	5	2.7394	1.3474	272	2.03	0.043
GRADE	8	0	.	.	.	.
Total_Subscale		0.2109	0.01642	273	12.85	<.0001
linear*Total_Subscale		0.0444	0.005314	274	8.35	<.0001

Table 23
HLM Estimates for MCA Predicting Common Reading Aloud

READING COMMON <- MCA

Effect	Grade	Estimate	SE	df	t	p
Intercept		6.1866	11.4446	366	0.54	0.5891
linear		2.3369	3.263	366	0.72	0.4743
grade	3	-94.1508	4.0503	368	-23.25	<.0001
grade	5	-51.3746	3.8353	368	-13.4	<.0001
grade	10	0	.	.	.	.
MCAread05		0.1199	0.007232	366	16.58	<.0001
linear*MCAread05		0.006815	0.002089	365	3.26	0.0012

Table 24
HLM Estimates for MCA Predicting Grade Level Reading Aloud

READING GRADE <- MCA

Effect	Grade	Estimate	SE	df	t	p
Intercept		-29.4431	10.3718	367	-2.84	0.0048
linear		5.4575	2.9797	364	1.83	0.0678
grade	3	-32.3119	3.6804	368	-8.78	<.0001
grade	5	0.8526	3.4853	368	0.24	0.8069
grade	10	0	.	.	.	.
MCAread05		0.1108	0.006554	366	16.91	<.0001
linear*MCAread05		0.004199	0.001908	363	2.2	0.0283

Table 24
HLM Estimates for MCA Predicting Common Maze (3 min)

MAZE COMMON <- MCA

Effect	Grade	Estimate	SE	df	t	p
Intercept		-6.1316	3.3473	368	-1.83	0.0678
linear		-3.1214	1.2216	365	-2.56	0.011
grade	3	-24.9609	1.194	368	-20.9	<.0001
grade	5	-16.4111	1.1307	368	-14.51	<.0001
grade	10	0	.	.	.	.
linear*grade	3	0.1912	0.4348	362	0.44	0.6605
linear*grade	5	1.2267	0.4139	364	2.96	0.0032
linear*grade	10	0	.	.	.	.
MCAread05		0.03205	0.002115	368	15.16	<.0001
linear*MCAread05		0.005522	0.000771	364	7.17	<.0001

Table 25
HLM Estimates for MCA Predicting Grade Level Maze (3 min)

MAZE GRADE <- MCA

Effect	Estimate	SE	df	t	p
Intercept	-19.4123	3.4353	370	-5.65	<.0001
linear	-1.5458	1.1051	369	-1.4	0.1627
MCAread05	0.03069	0.002203	370	13.93	<.0001
linear*MCAread05	0.004744	0.000708	368	6.71	<.0001

Figure Caption

Figure 1. Seamless and flexible system of progress monitoring.

Figure 2. Alternate-form reliability for reading aloud and maze selection across time and grade levels.

Figure 3. Correlations between reading aloud and maze selection and the NALT/MAP across grade levels.

Figure 4. Linearity of relation between CBM and NALT/MAP measures.

Figure 5. Correlations between reading aloud and maze selection and the MCAs across grade levels.

Figure 6. Linearity of relation between CBM and MCA measures.

Figure 7. Change in reading aloud scores from Fall to Winter to Spring.

Figure 8. Change in common maze selection scores from Fall to Winter to Spring.

Figure 9. Change in grade level maze selection scores from Fall to Winter to Spring.

Figure 10. Fall across grade growth for reading aloud and maze selection.

Figure 11. Total NALT/MAP subscale predicting common reading aloud.

Figure 12. Total NALT/MAP subscale predicting common maze selection.

Figure 13. Total NALT/MAP subscale predicting grade level maze selection.

Figure 14. MCA predicting common reading aloud.

Figure 15. MCA predicting grade level reading aloud.

Figure 16. MCA predicting common maze selection (3 min).

Figure 17. MCA predicting grade level maze selection (3 min).

Figure 1. Seamless and flexible system of progress monitoring.

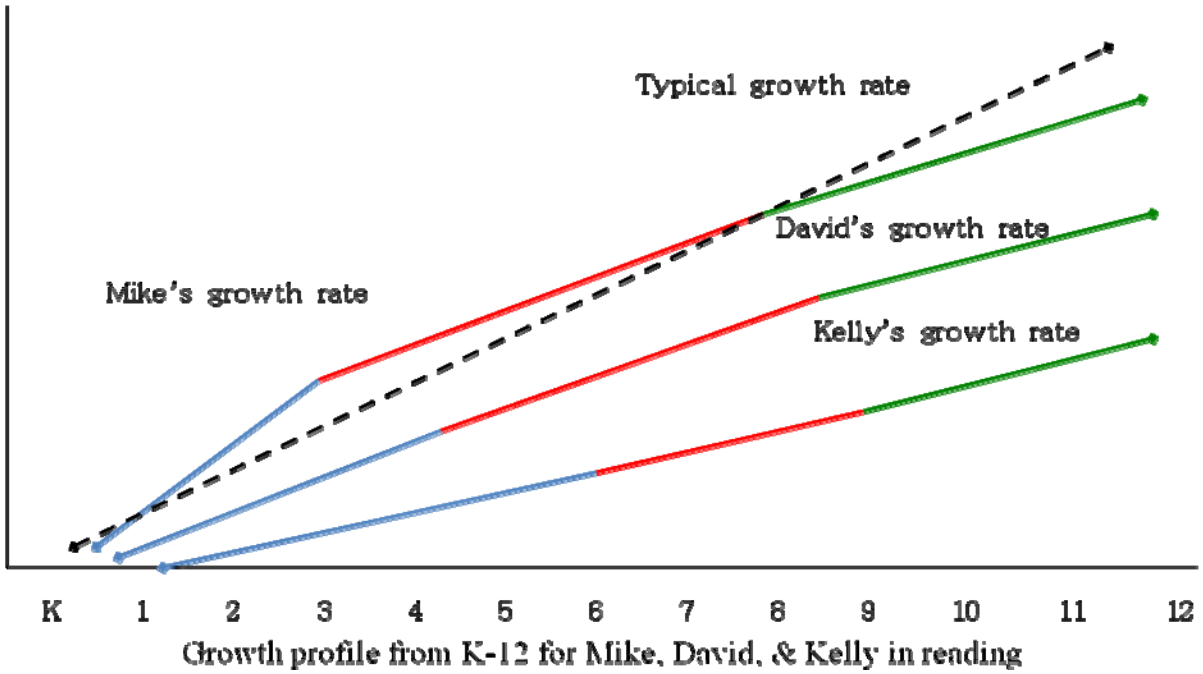


Figure 2. Alternate-form reliability for reading aloud and maze selection across time and grade levels.

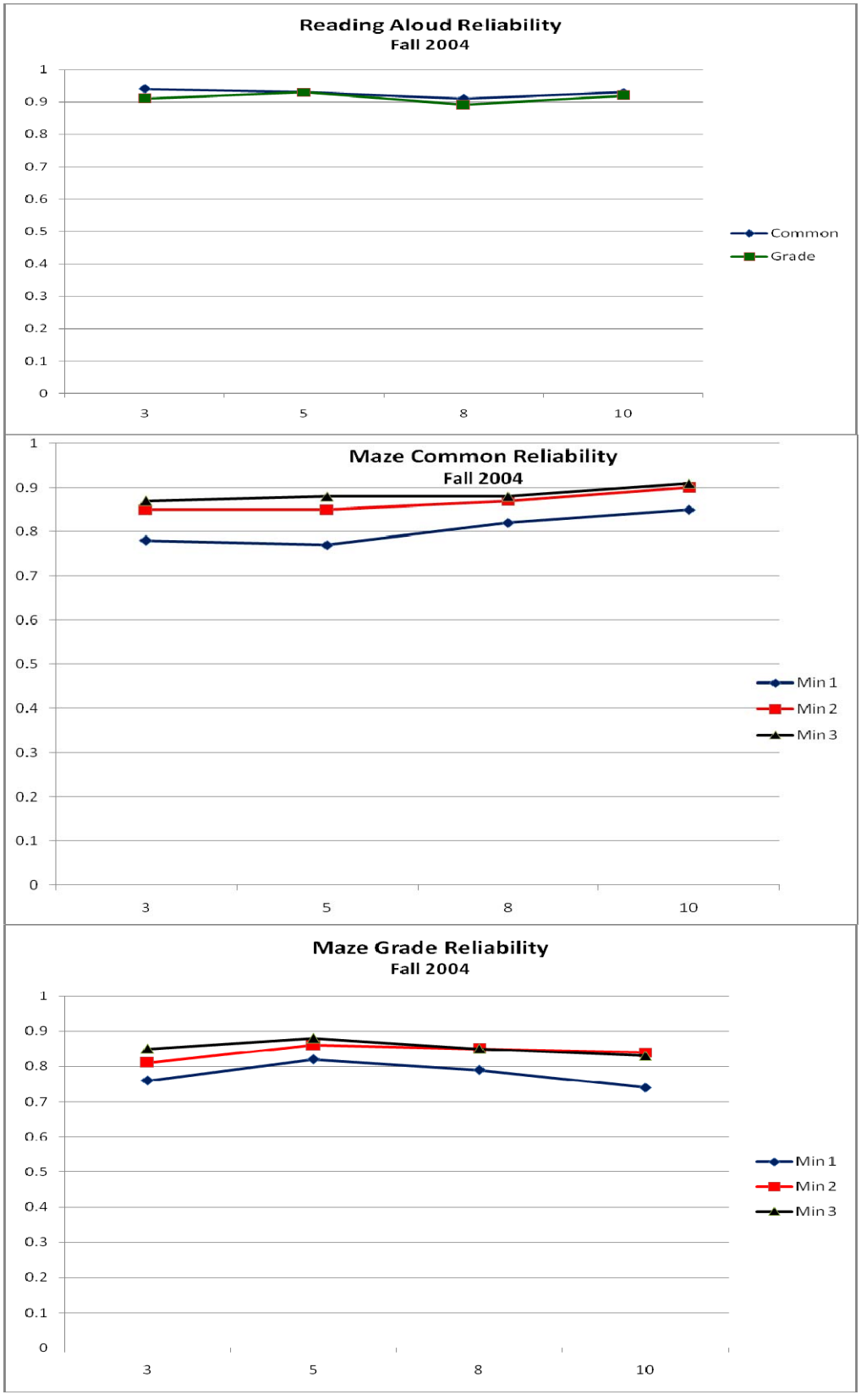


Figure 2 (continued).

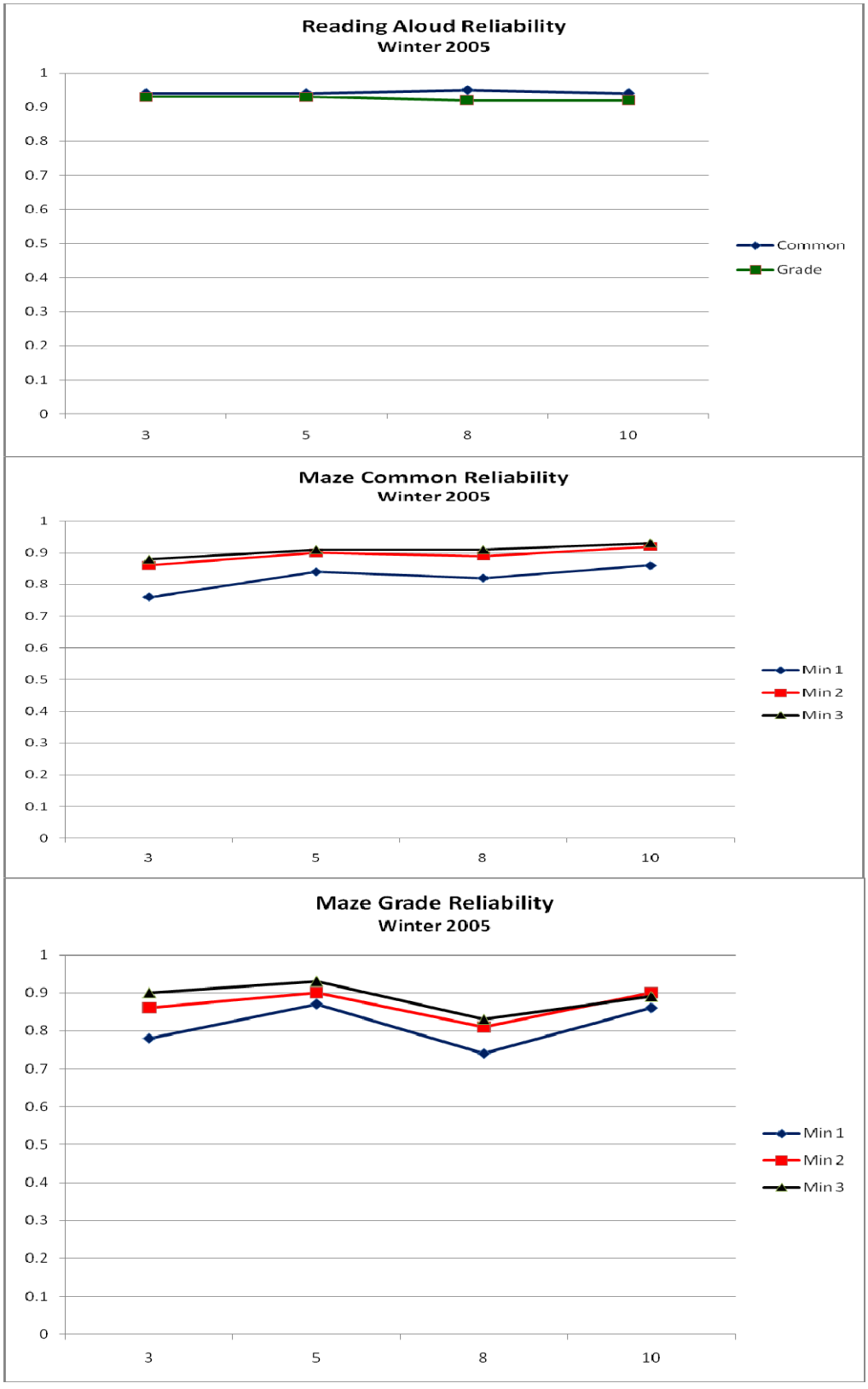
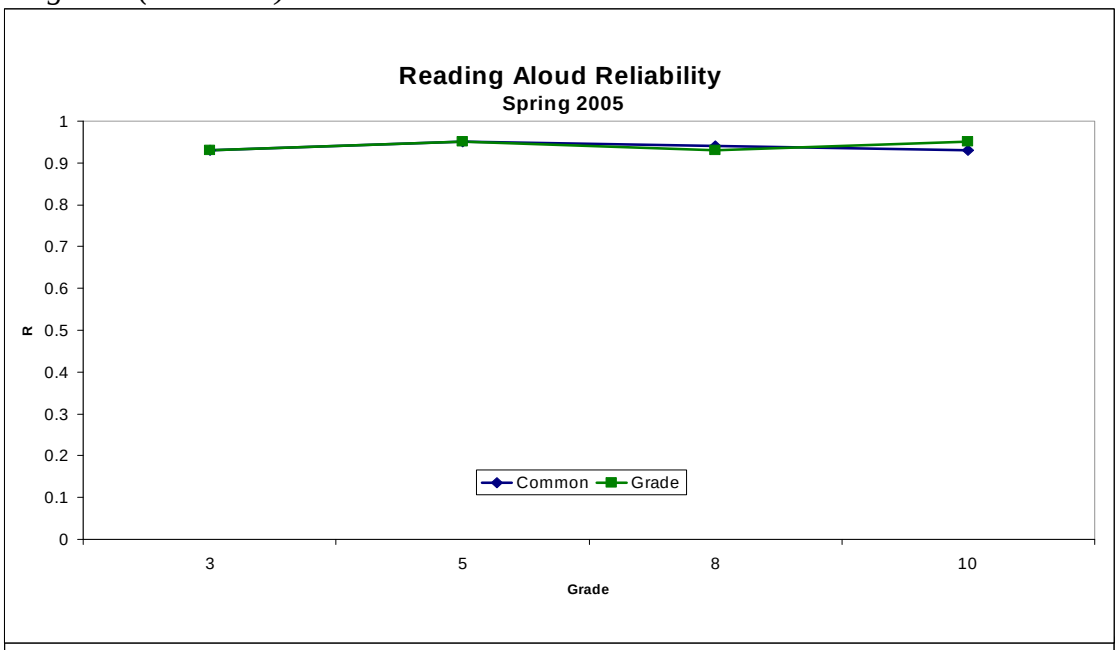
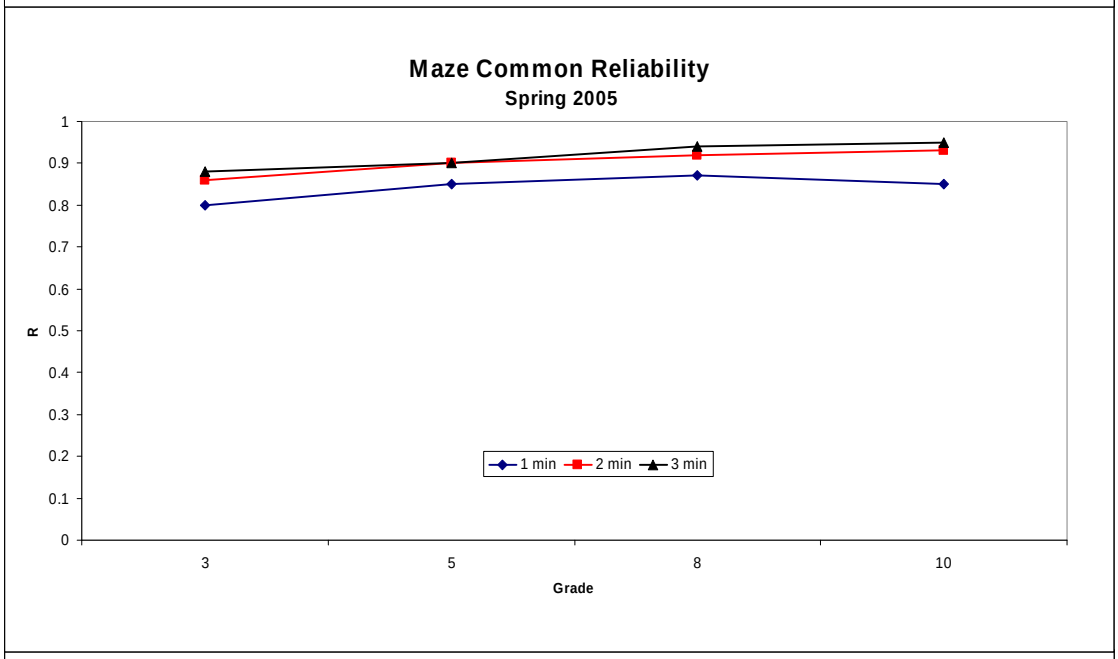
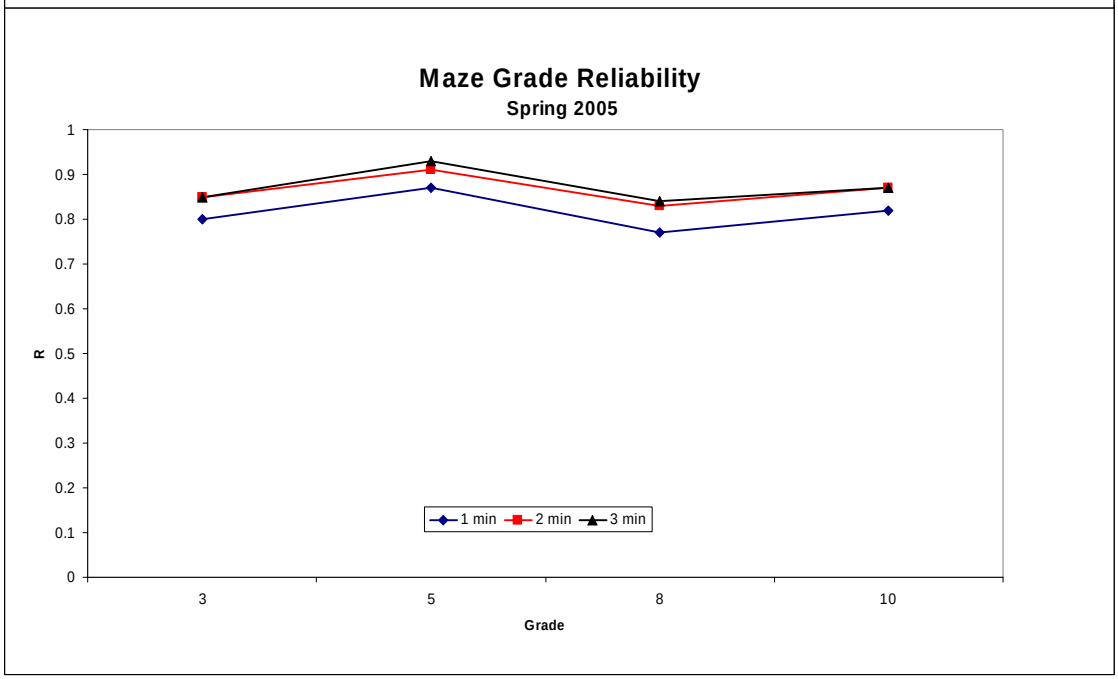
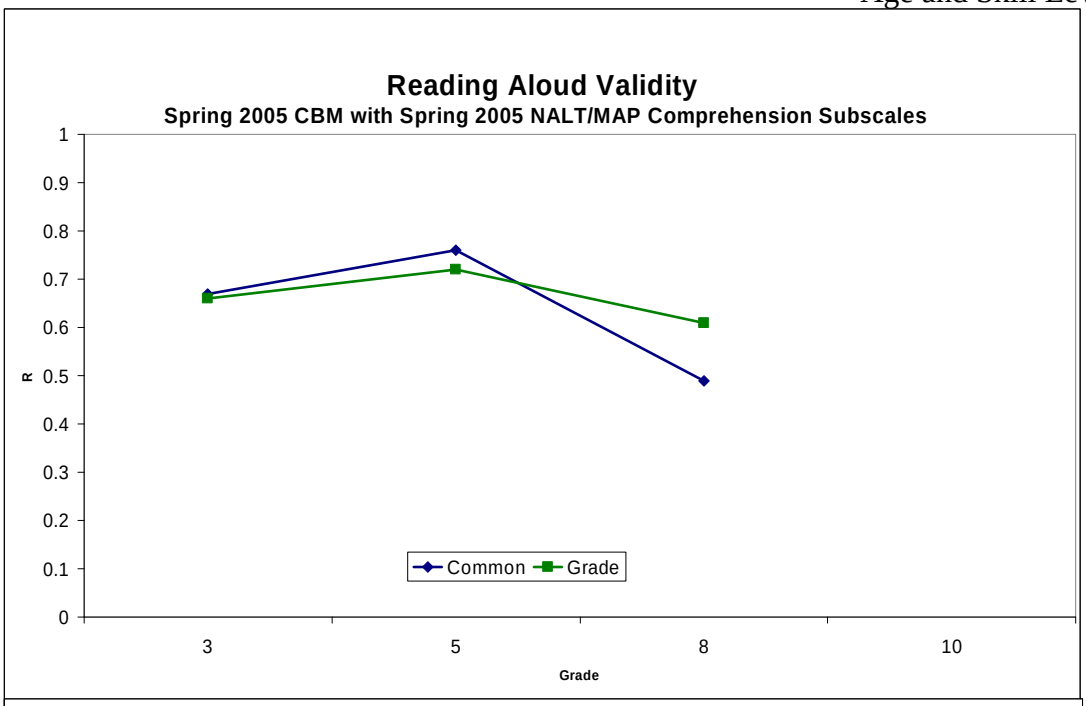


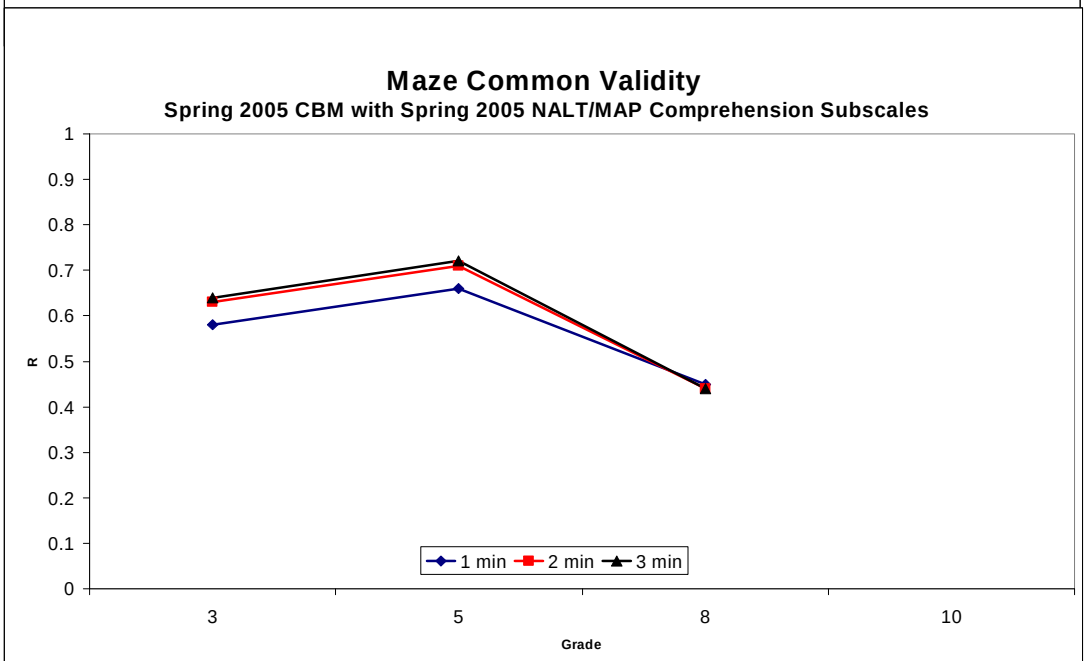
Figure 2 (continued).











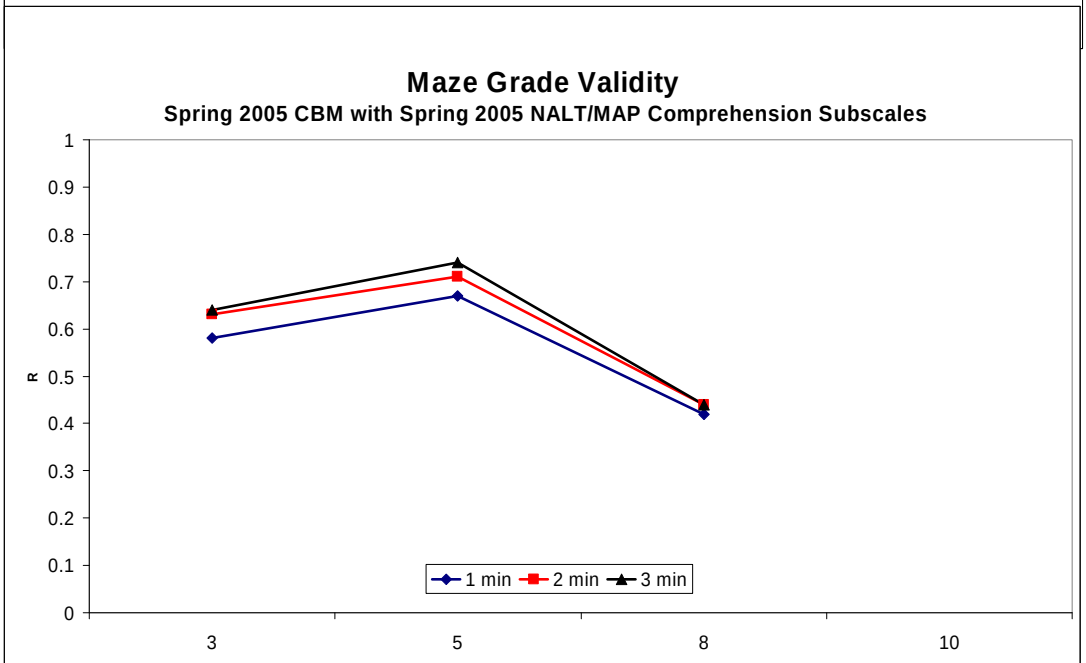


Figure 3. Correlations between reading aloud and maze selection and the NALT/MAP across grade levels.

Figure 4. Linearity of relation between CBM and NALT/MAP measures.

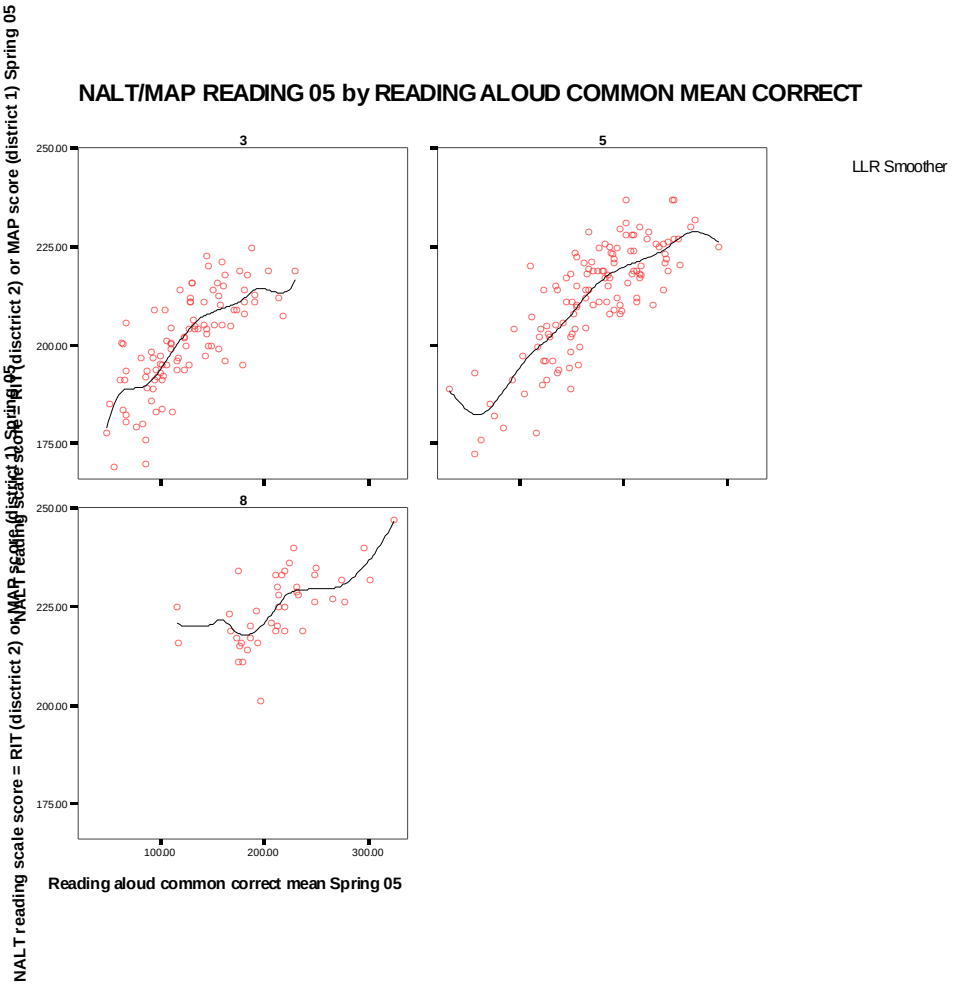


Figure 4(continued).

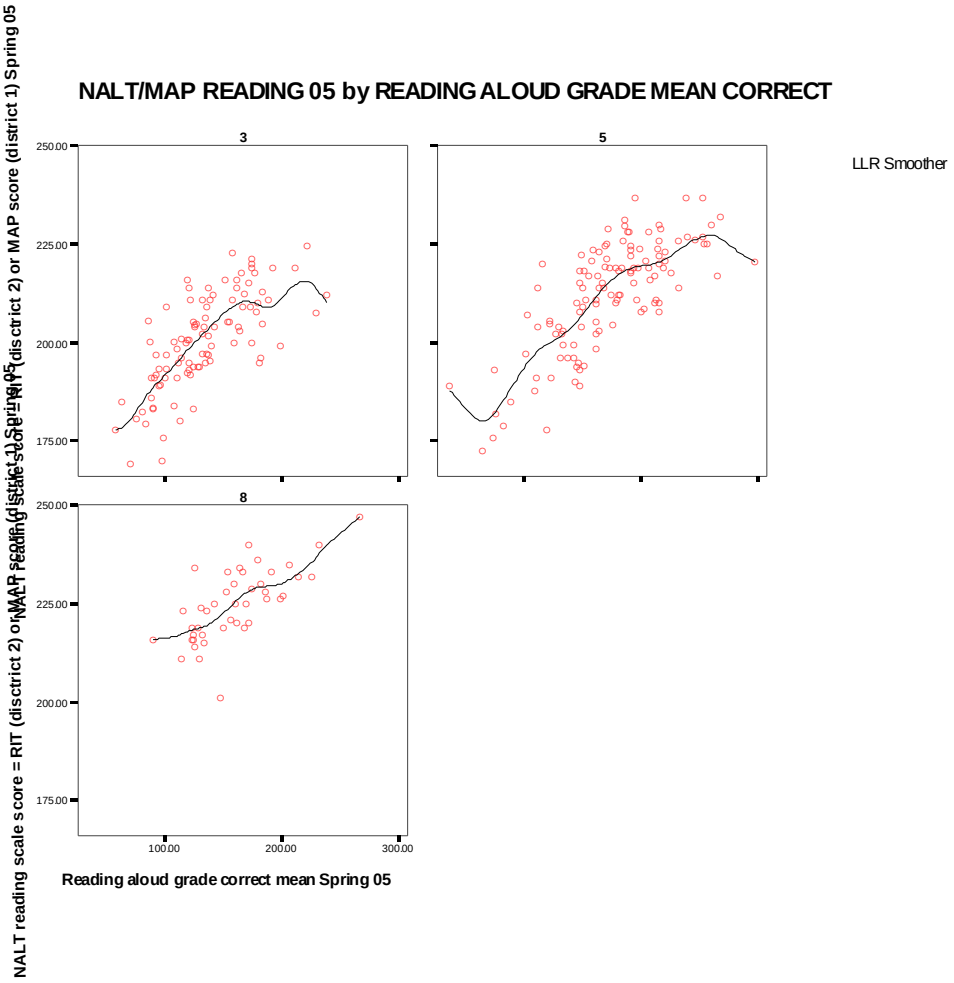


Figure 4(continued).

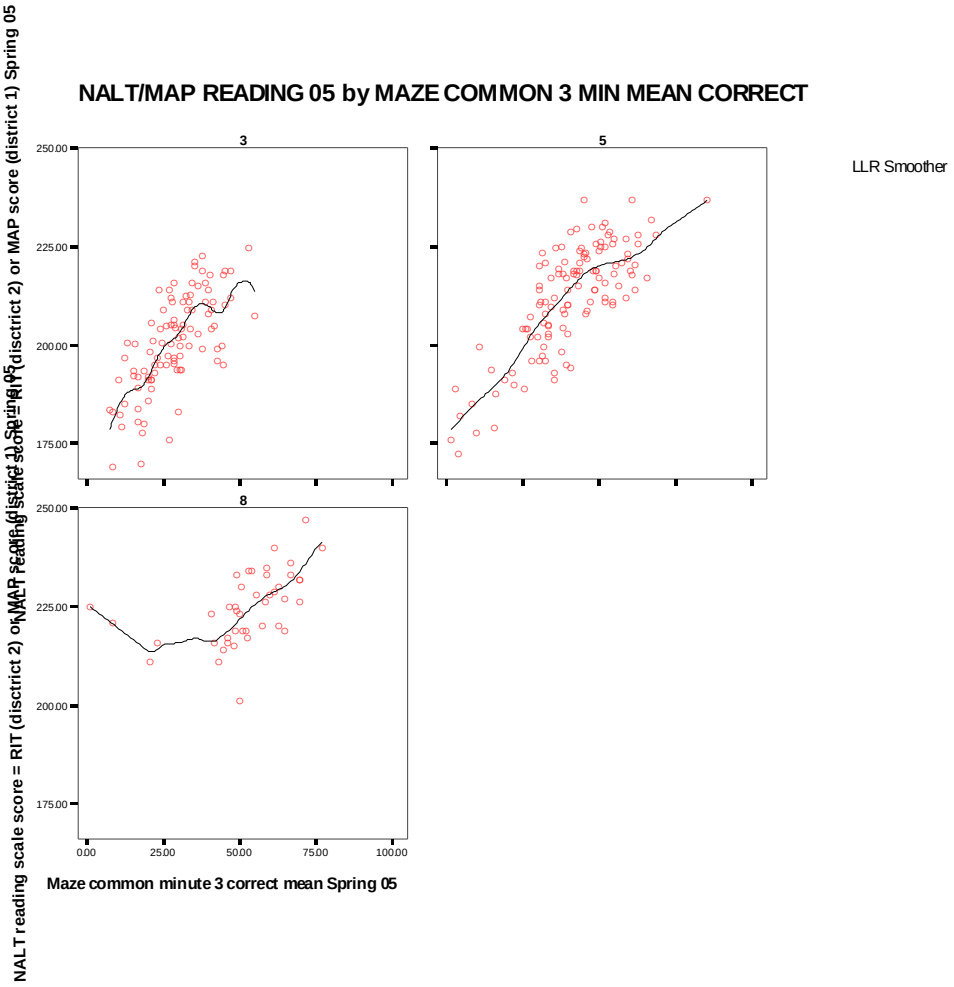
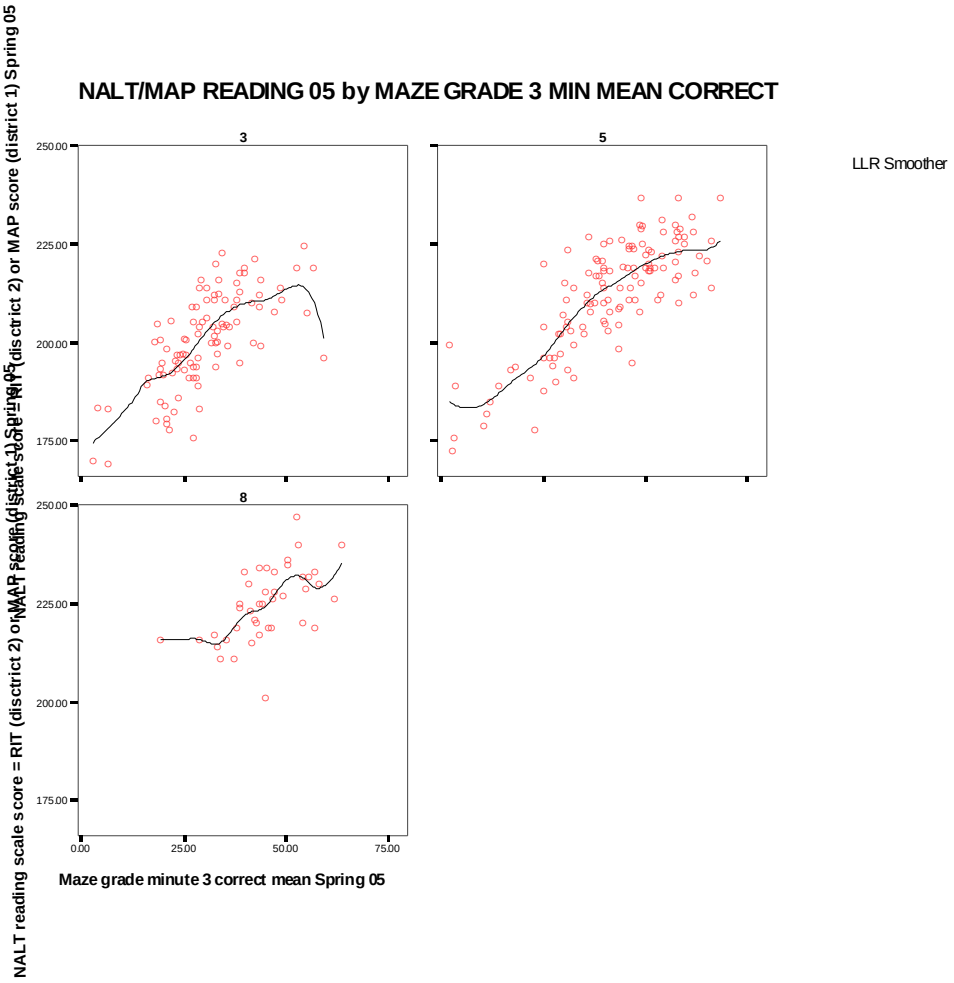
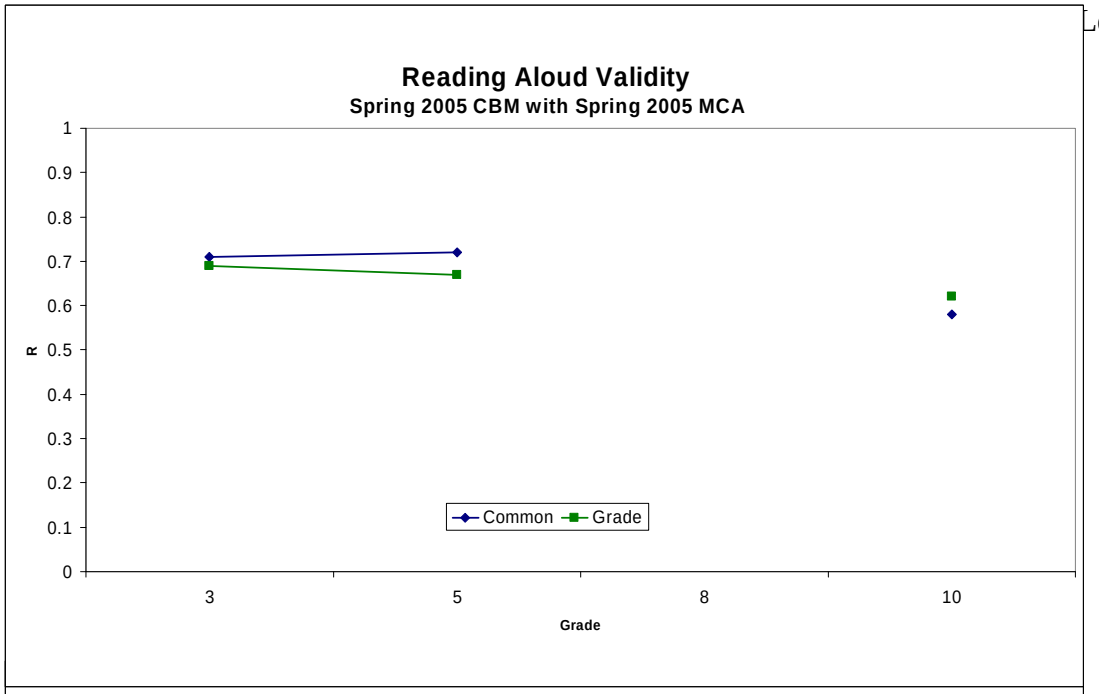
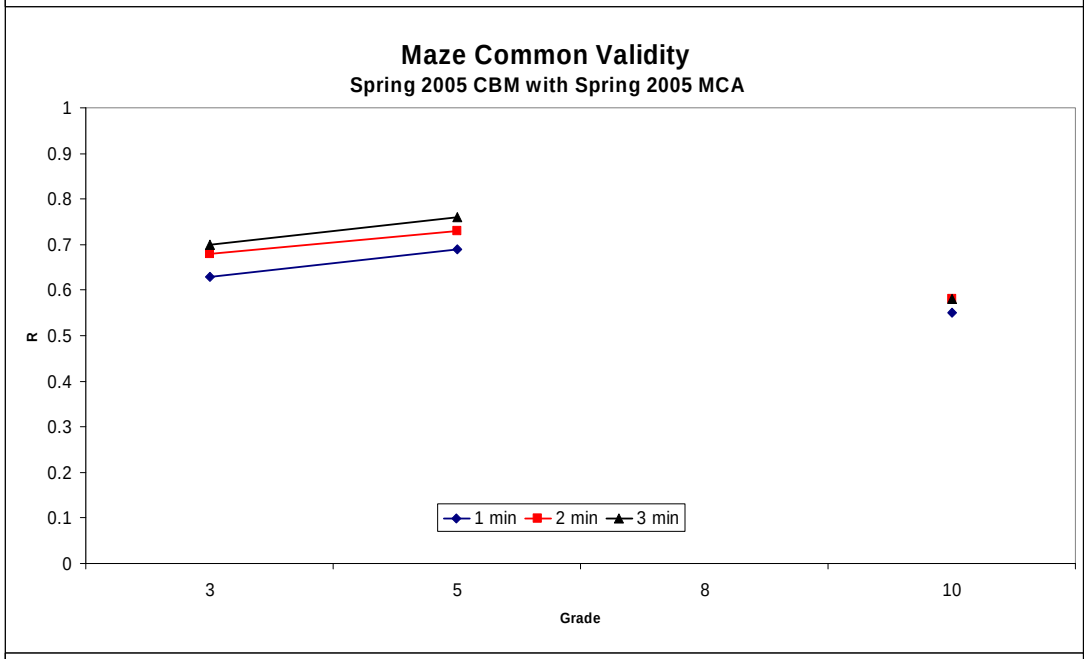


Figure 4 (continued).







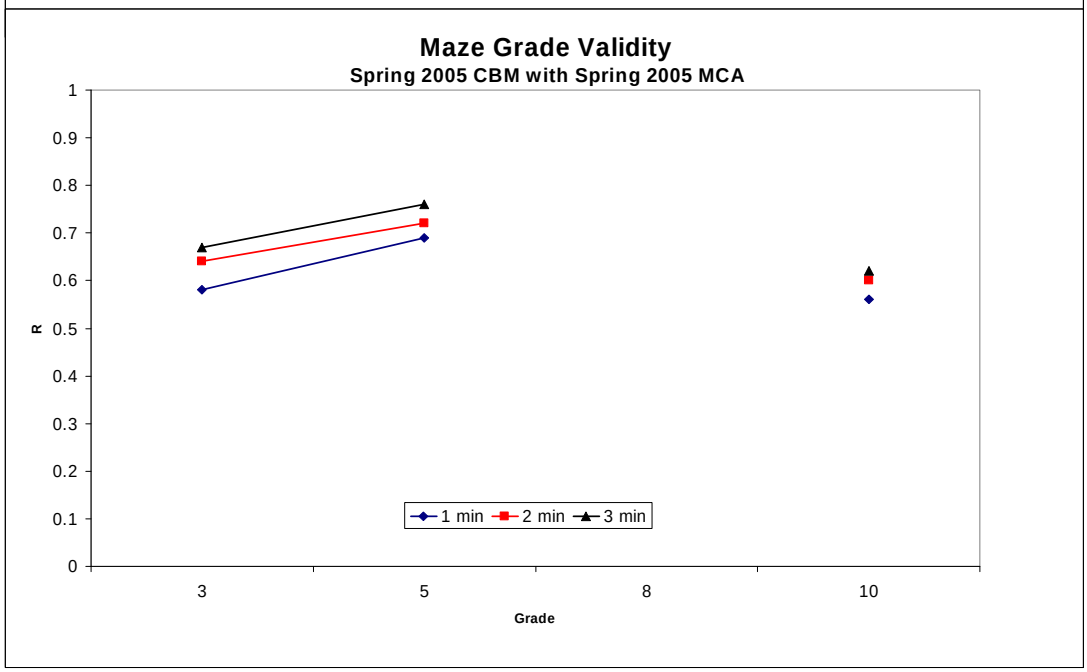


Figure 5. Correlations between reading aloud and maze selection and the MCAs across grade levels.

Figure 6. Linearity of relation between CBM and MCA measures.

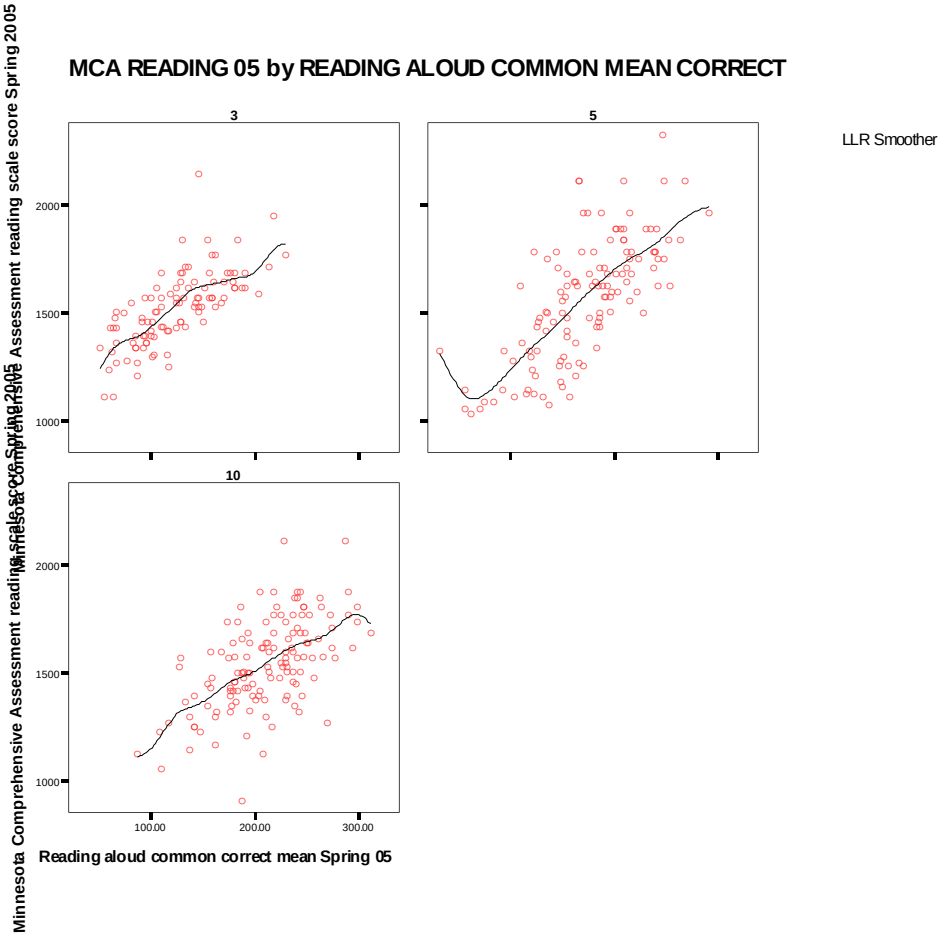


Figure 6 (continued).

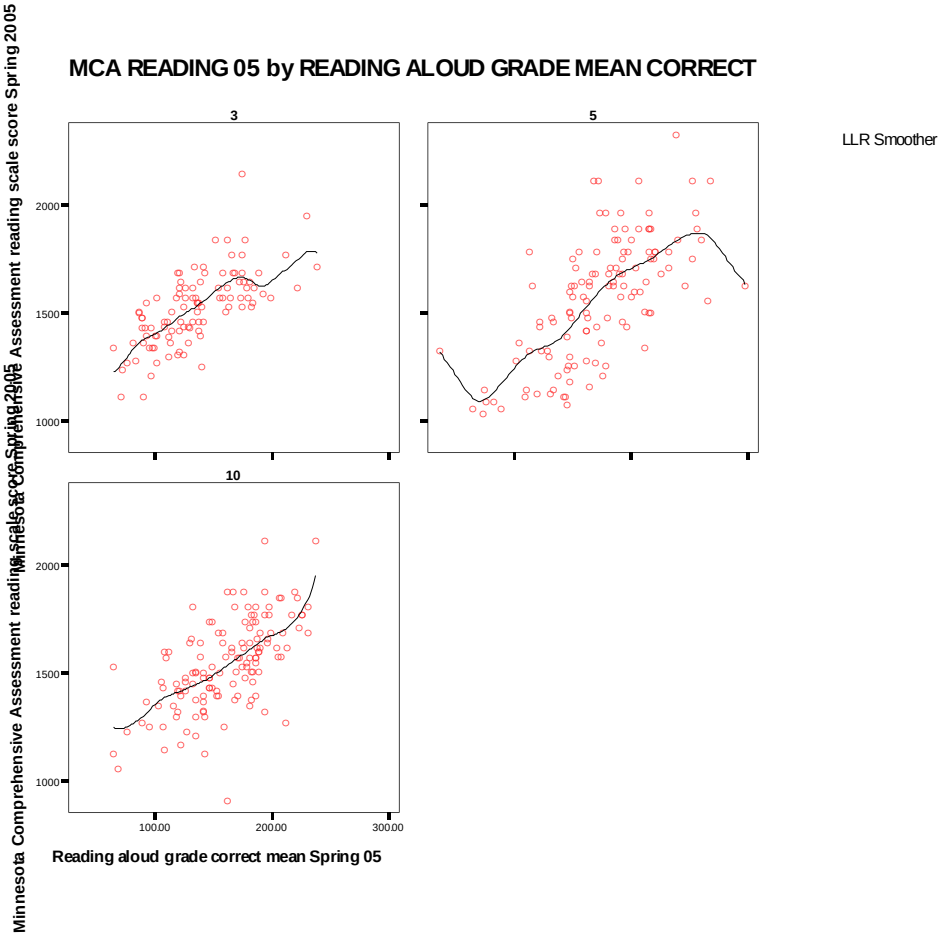


Figure 6 (continued).

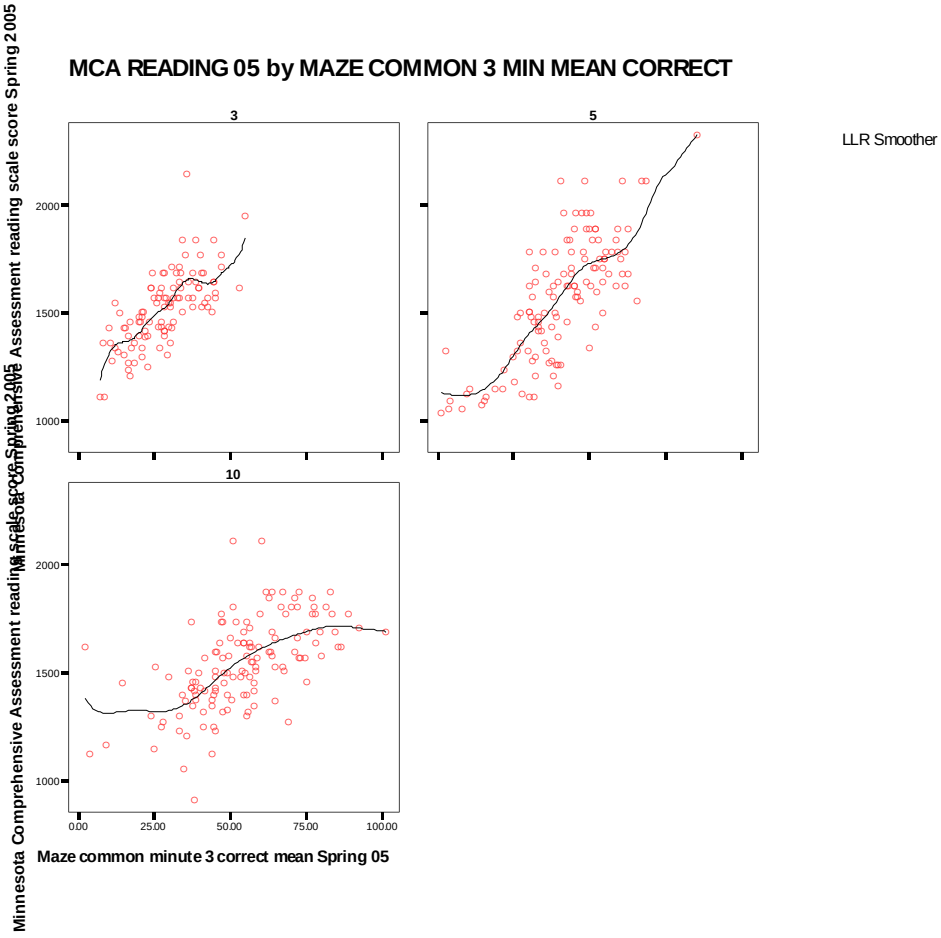
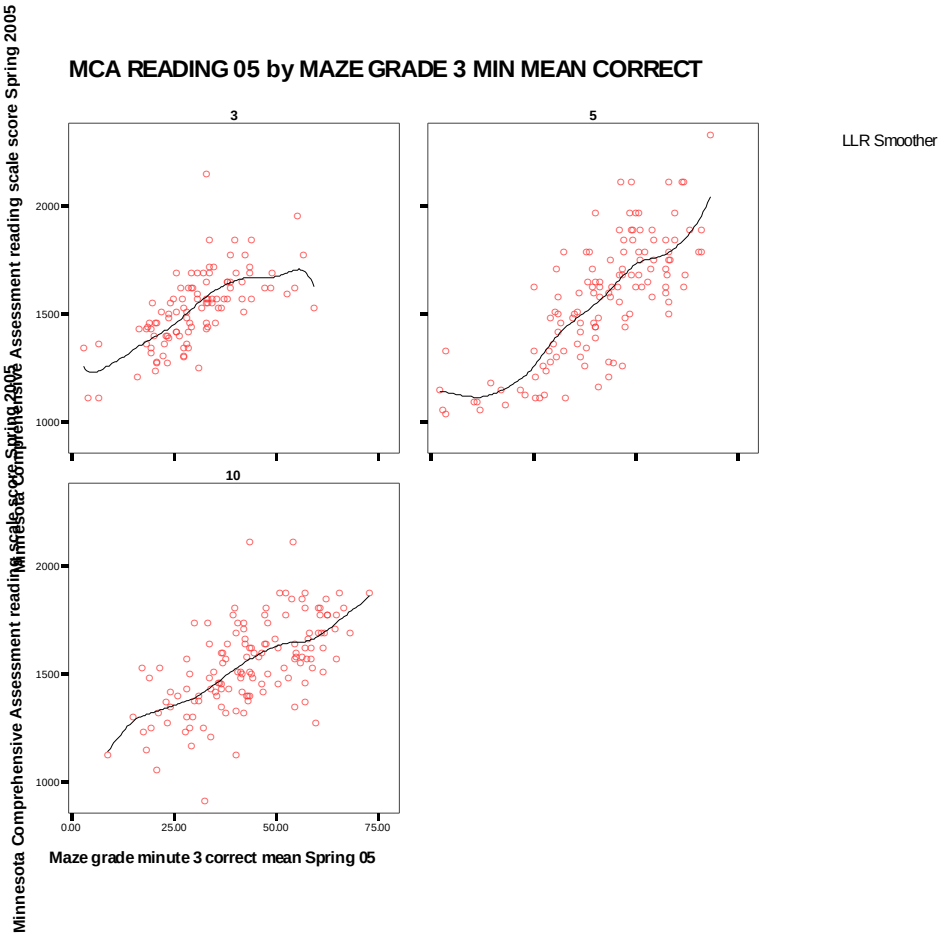


Figure 6 (continued).



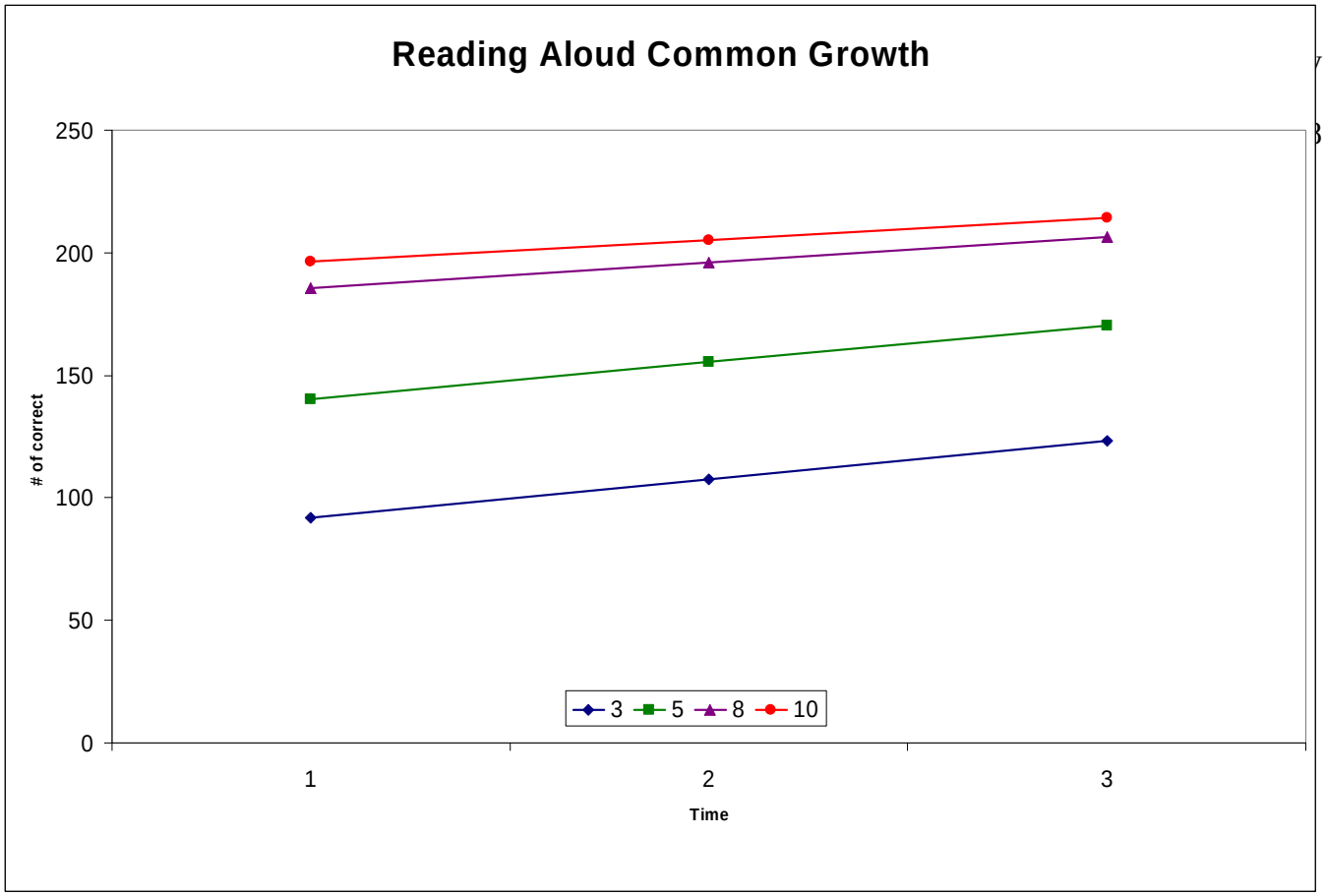
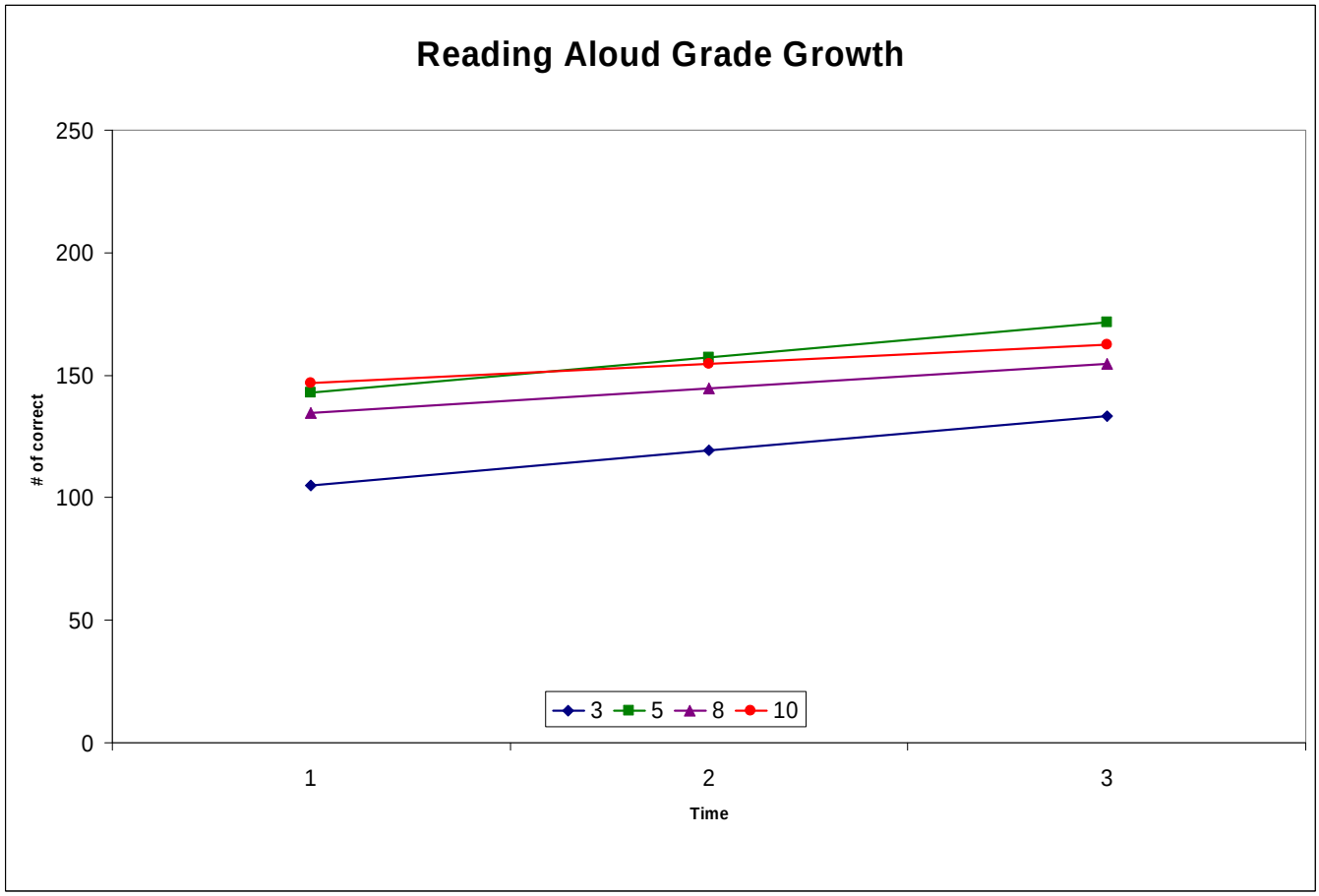
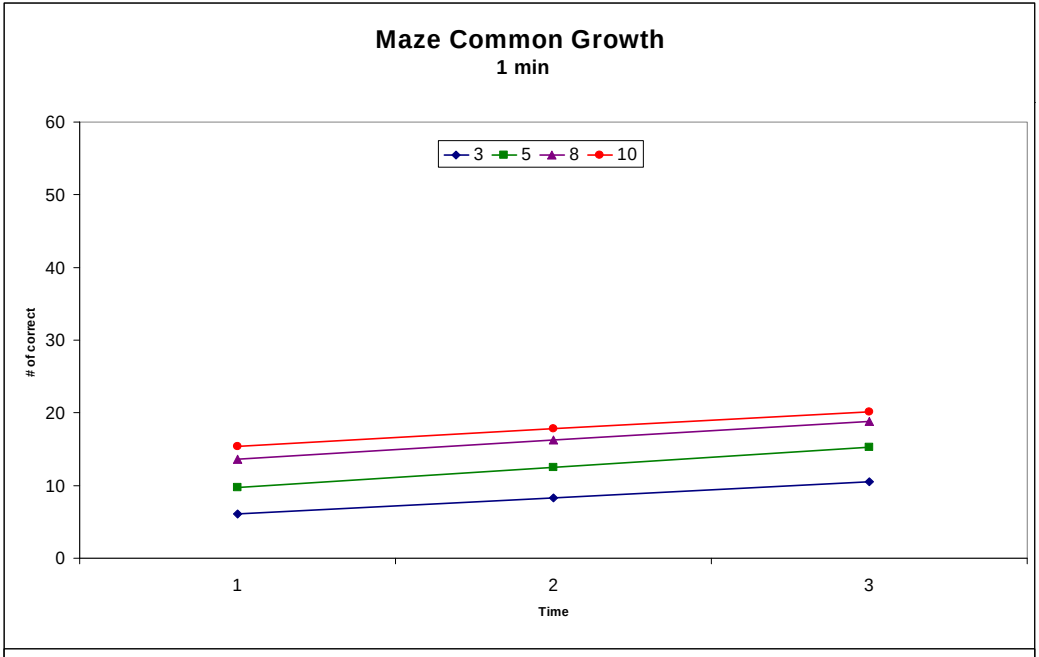
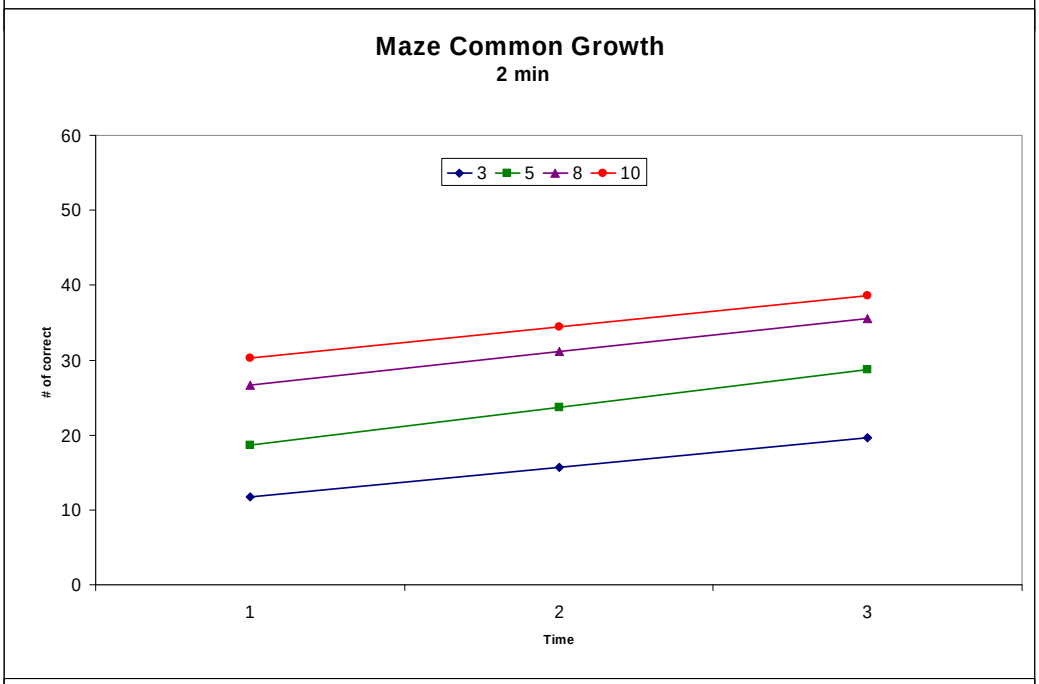


Figure 7. Change in reading aloud scores from Fall to Winter to Spring.







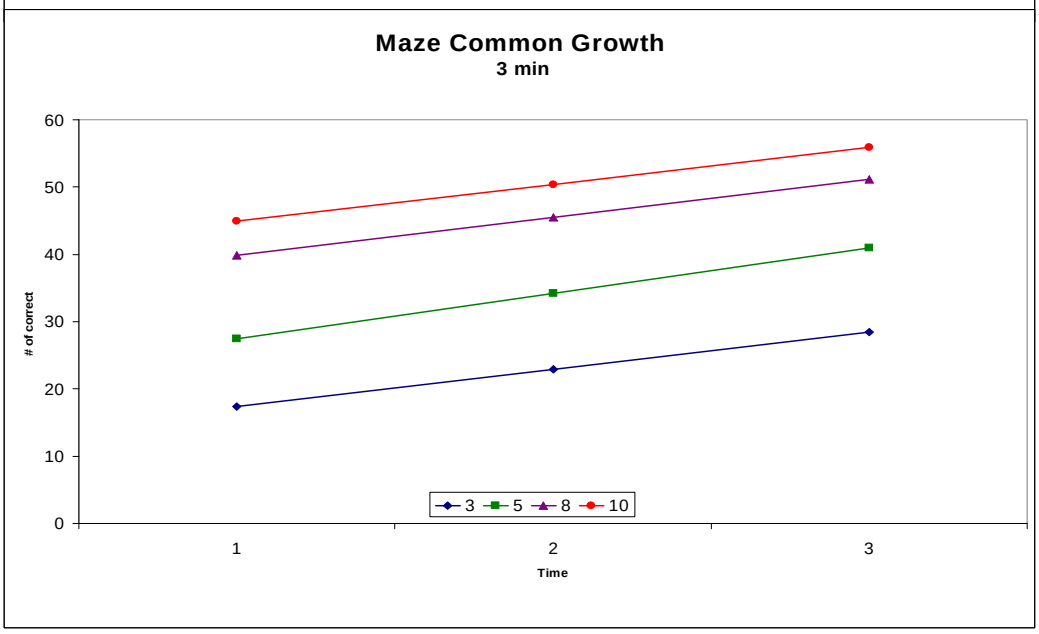
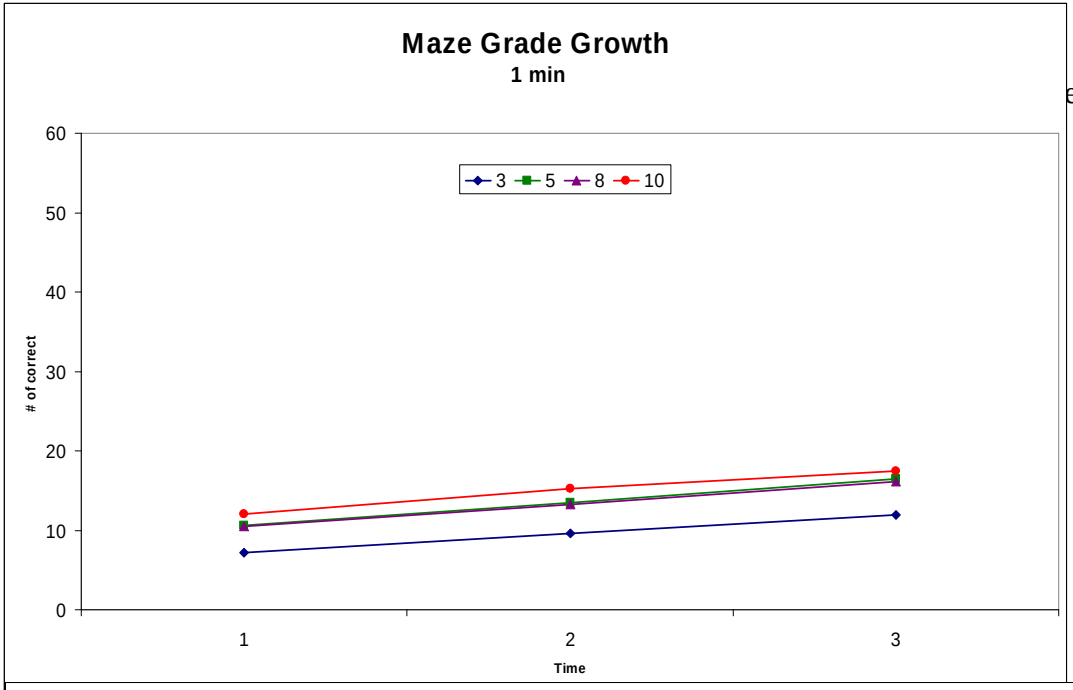
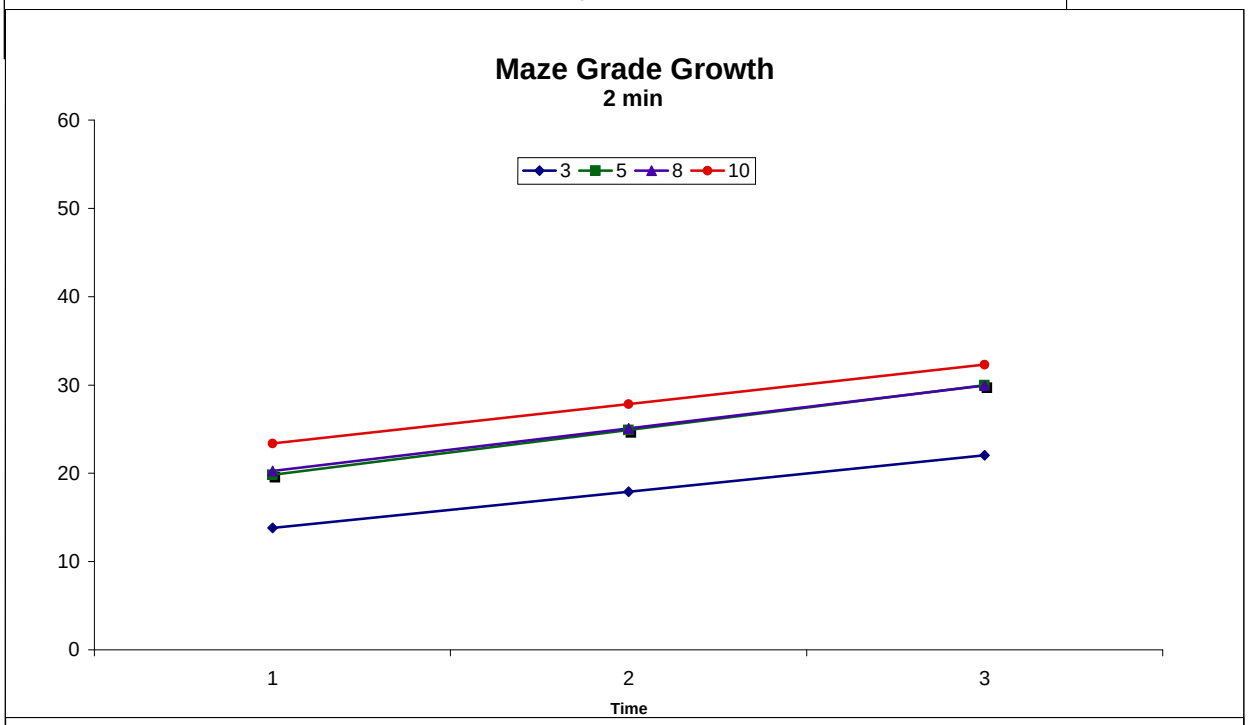


Figure 8. Change in common maze selection scores from Fall to Winter to Spring.





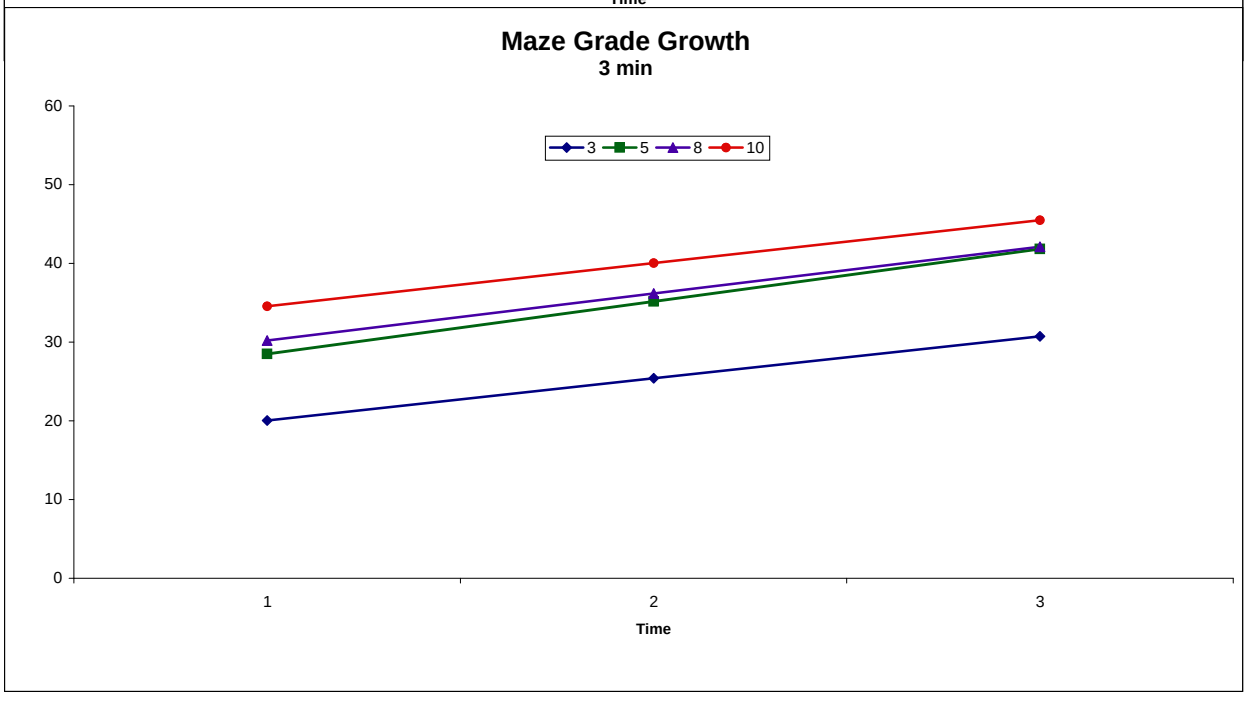


Figure 9. Change in grade level maze selection scores from Fall to Winter to Spring.

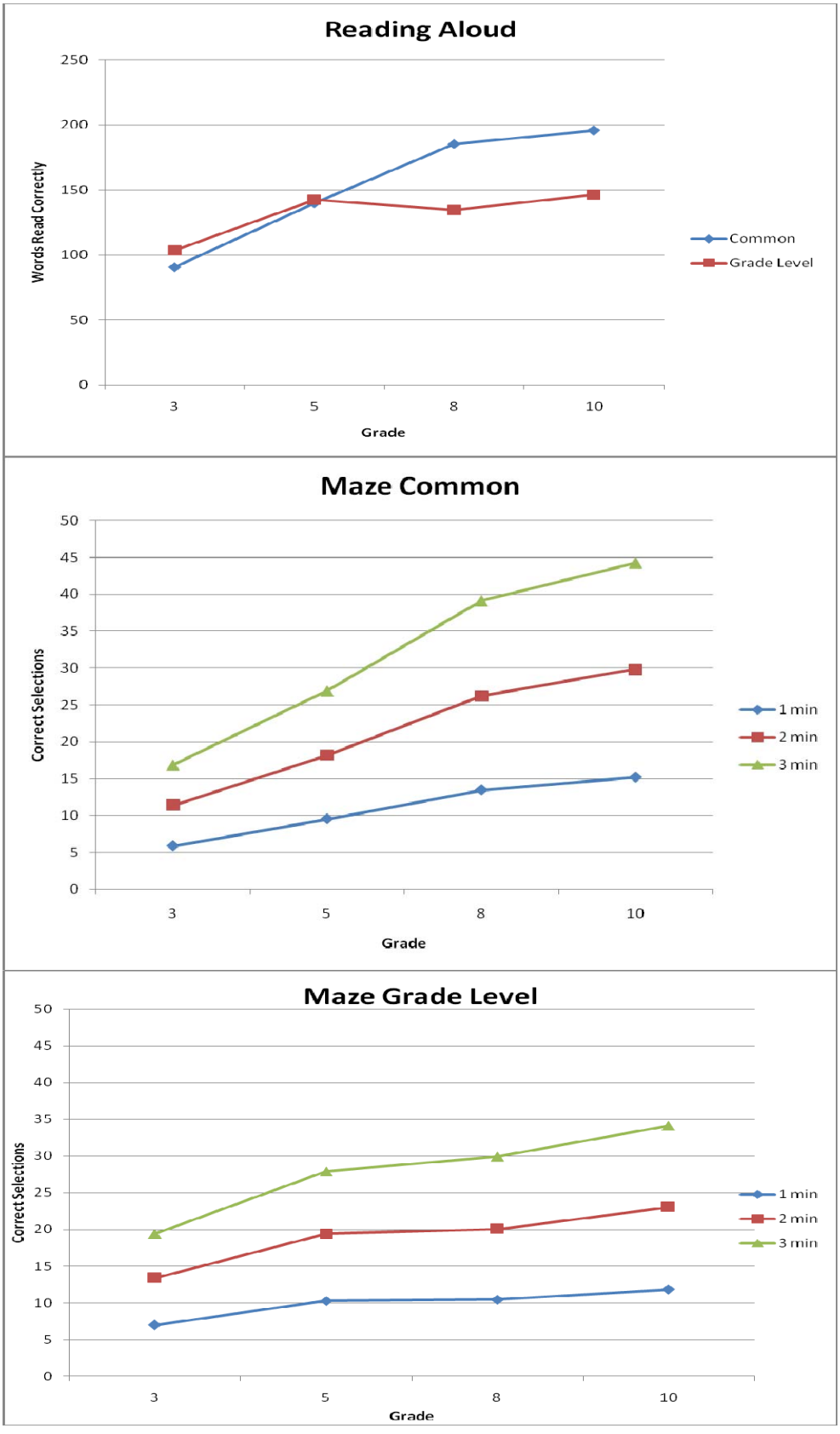


Figure 10. Fall across grade growth for reading aloud and maze selection.

Figure 11. Total NALT/MAP subscale predicting common reading aloud.

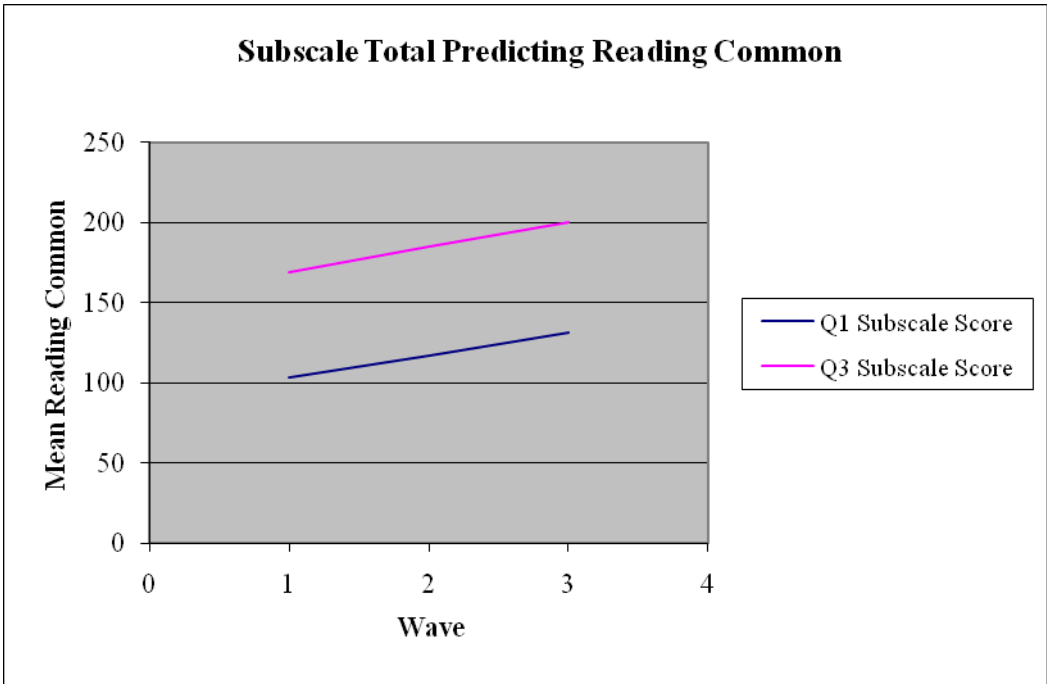


Figure 12. Total NALT/MAP subscale predicting common maze selection.

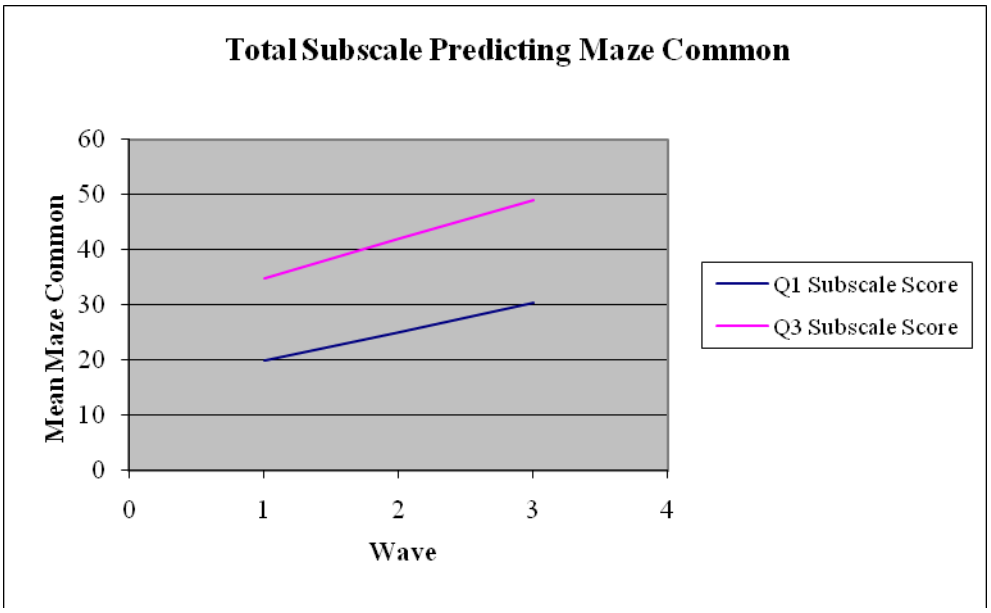


Figure 13. Total NALT/MAP subscale predicting grade level maze selection.

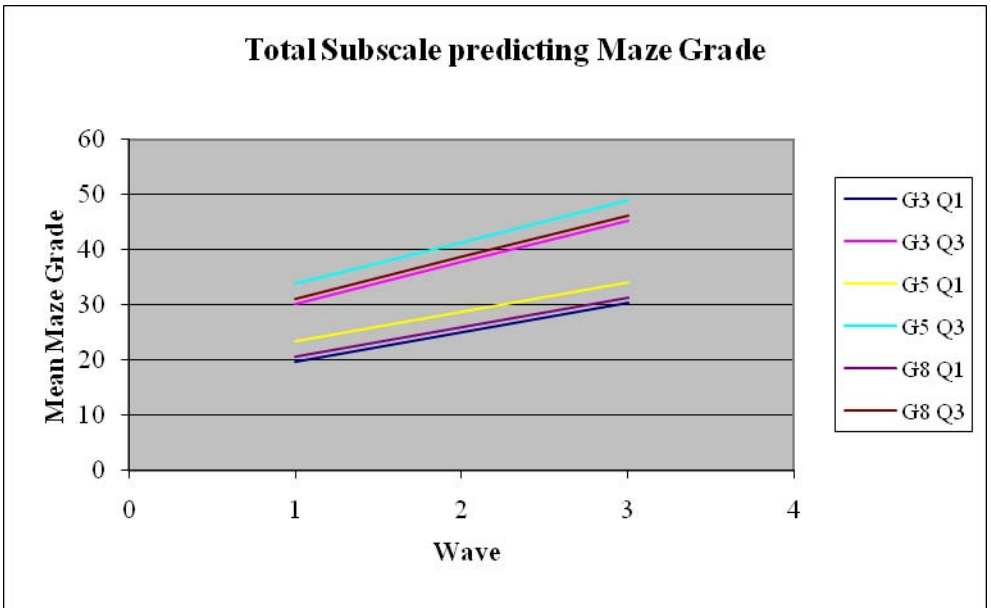


Figure 14. MCA predicting common reading aloud.

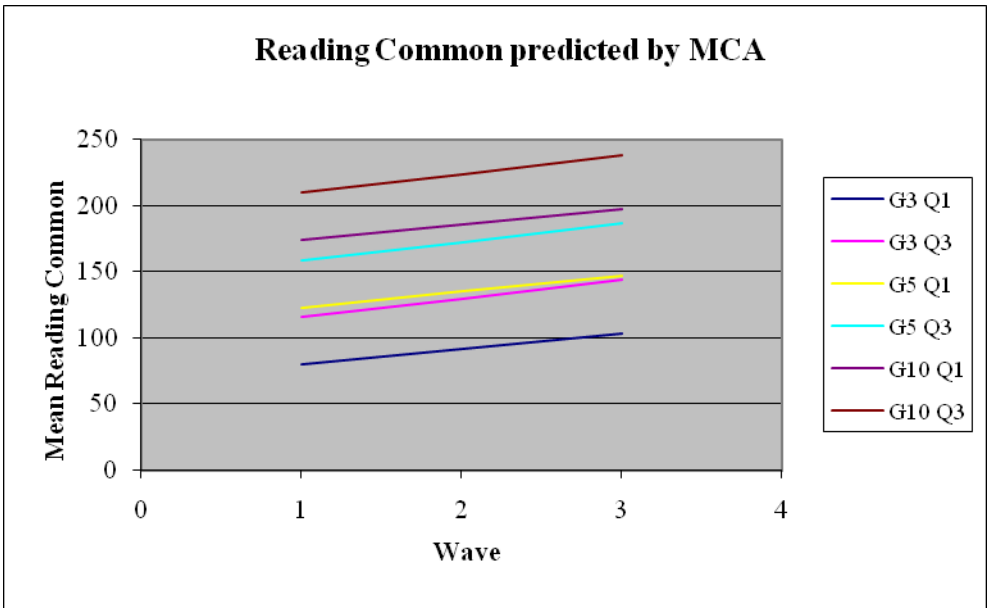


Figure 15. MCA predicting grade level reading aloud.

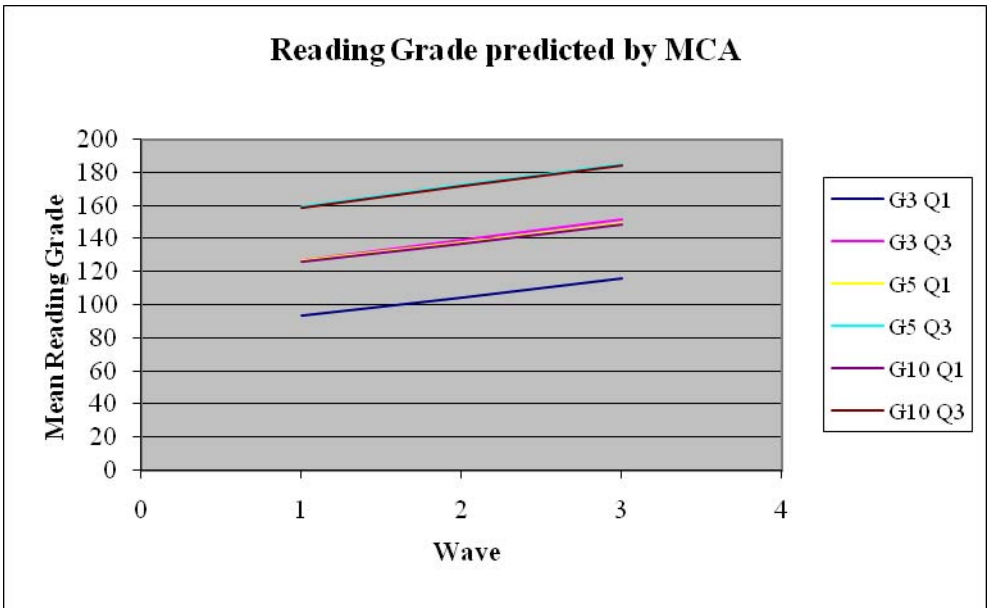


Figure 16. MCA predicting common maze selection (3 min).

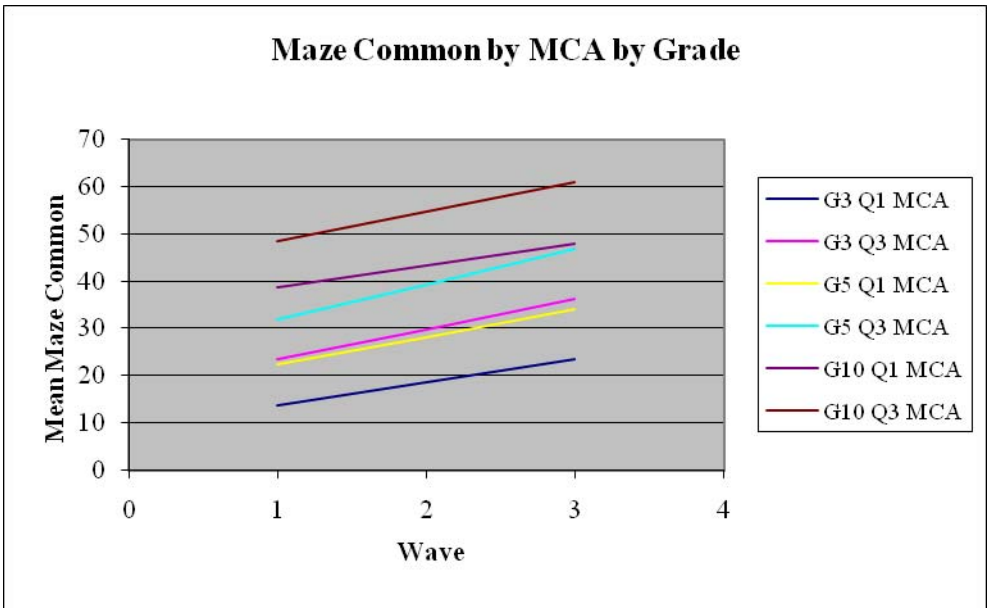


Figure 17. MCA predicting grade level maze selection (3 min).