TECHNICAL REPORT #27:

Study of General Outcome Measurement (GOMs) in Reading for Students with Significant Cognitive Disabilities: Year 1

Teri Wallace, Renáta Tichá, and Kathy Gustafson

RIPM Year 4: 2006 – 2007

Dates of Study: October 2006 – June 2007

March 2008

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.

Abstract

The goal of this study was to examine the technical characteristics of newly created general outcome measures (GOMs) in reading for students with significant cognitive disabilities. The participants in the study were 31 students with significant cognitive disabilities. It was found that the GOMs used in this study produced reliable data. The results establishing validity of the GOMs suggest that Rimes, K-4 Science Content Word Identification, Simple Maze, and Fry's 100 Sight Word Identification GOMs are the most promising measures in predicting student reading performance. Functional Pictures and Signs, Mixed Letter Identification, Fry's 100 Sight Word Identification, and K-4 Science Content Word Identification showed the best potential for measuring reading progress of students with significant cognitive disabilities. The results suggest GOMs to be a useful tool for teachers to predict and potentially monitor progress of students with significant cognitive disabilities in reading. More research is needed to replicate these results.

Study of General Outcome Measurement (GOMs) in Reading for Students with Significant Cognitive Disabilities: Year 1

Federal Requirements for Success and Assessment

The No Child Left Behind Act (NCLB) of 2001 provides a legal mandate to ensure that all students are learning. This law requires the development of state standards and large scale assessments intended to measure schools' success in achieving established content and achievement standards, including standards in areas related to literacy. This legislation (e.g., Public Law No. 107-110, 115 Stat. 1425, 2002) has mandated that students with significant cognitive disabilities be included in states' accountability systems.

IDEA regulations, published in the Federal Register in December 2003, provide an avenue for students with disabilities to be assessed through one of five options as determined by the child's IEP team, including:

- The regular grade-level state assessment,
- The regular grade-level state assessment with accommodations,
- Alternate assessments aligned with grade-level achievement standards,
- Alternate assessments based on alternate achievement standards, or
- Modified achievement standards.

Alternate assessments, in general, are intended for use with students with disabilities who are unable to participate meaningfully in general state and district assessment systems, even with accommodations (Roach & Elliott, 2006). Alternate assessments based on alternate achievement standards are for students with significant cognitive disabilities who cannot meet typical grade-level achievement standards.

In 2002, NCLB increased the federal government's emphasis on assessment and accountability systems. Noted by Roach and Elliott (2006), many states have struggled to develop alternate assessments that meet federal mandates for students with significant cognitive disabilities for two primary reasons. First, the skills and concepts in the state academic standards were considered inappropriate or irrelevant for students with significant cognitive disabilities, which resulted in alternate assessments that focused on functional domains; and second, the development of alternate assessments was considered a special education function and deemed to be only somewhat connected to states' overall assessment systems. However, the reauthorization of IDEA (2004) and guidelines for using alternate assessment with alternate achievement standards for NCLB (Federal Register, December 9, 2003) both require determining adequate yearly progress for this population using alternate assessments that are linked to the state’s academic content standards. States may use alternate achievement standards for up to 1% of students with significant cognitive disabilities and modified achievement standards for up to 2% of students with persistent academic difficulties.

Assessing the academic performance and progress of students with significant cognitive disabilities has long been a challenge to the field of education (Browder & Spooner, 2006). The information gained from using standardized tests with students with significant cognitive disabilities may not provide useful information for teachers to use in educational decision-making (Allinder & Siegel, 1999; Duncan, Sbardellati, Maheady, & Sainato, 1981; Sigafoos, Cole, & McQuarter, 1987). Additional assessment strategies may supplement standardized tests, including: criterion-referenced tests, adaptive behavior scales, observations, fluency measures, interviews and others. These may provide useful data for educational decision-making.

However, there is presently no consensus regarding the progress students with significant cognitive disabilities should make within the general education curriculum, including performance and progress on state standards and assessments. While portfolios and mastery monitoring strategies have been used as alternate assessments, each has its limitations. General outcome measurement (GOM) may be a useful solution as there is substantial evidence for its use with students with learning disabilities, to show both performance and growth as well as to predict future performance (see Marston, 1989; Stecker, Fuchs, & Fuchs, 2005; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). General outcome measurement (GOM) has roots in curriculum-based measurement (CBM; Deno, 1995).

Curriculum-Based Measurement

Curriculum-Based Measurement (CBM) is a method for monitoring student growth in an academic area and evaluating the effects of instructional programs on that growth (Deno, 1985). CBM is designed to be part of a problem-solving approach to special education where the academic difficulties of students were viewed as problems to be solved rather than as unchangeable characteristics within a child (Deno, 1990). In the problem-solving approach, teachers are the “problem solvers” who constantly evaluate and modify students’ instructional programs. For a problem-solving approach to be effective, it is necessary for teachers to have a tool that can be used to evaluate growth in response to instruction. CBM was developed to serve that purpose.

Two separate but related concerns drove the initial research in the development of CBM (Deno, 1985). The first was the concern for technical adequacy. If teachers were to use the measures to make instructional decisions, the measures would have to have demonstrated reliability and validity. The second was the concern for practicality. If teachers were to use the measures on an ongoing and frequent basis to evaluate instructional programs, the measures would have to be simple, efficient, easily understood, and inexpensive. Therefore, CBM measures were conceptualized to be short samples of work that would be indicators of academic performance. The samples would need to be valid and reliable with respect to the broader academic domain they were representing. When the samples are taken from the curriculum used in the classroom, the term curriculum-based measurement (CBM) is used. In contrast, when the samples are taken from other materials, e.g. newspaper articles, the term general outcome measurement (GOM) is more appropriate (Deno, 2003).

Curriculum-based measurement has a 30-year research base establishing its reputation as an evidence-based practice in measuring individual performance and progress. CBM's most extensive research history is in reading for elementary-aged students, but there is also research in other instructional areas, such as: writing, spelling, math, science, and more (Allinder & Swain, 1997; Calhoon & Fuch, 2003; Espin & Deno, 1994; Espin et al., 2000; Espin et al., 2005; Foegen & Deno, 2001; Fuchs & Fuchs, 2002; Shin, Deno, & Espin, 2000). While CBM was originally intended for school-age students, it is now used with children across the age spectrum, pre-K through high school. Researchers have utilized the principles of CBM to create other progress monitoring systems, e.g. Individual Growth and Development Indicators (IGDIs) and Dynamic Indicators of Basic Literacy Skills (DIBELS) for children in preschool and daycare settings (Good & Kaminski, 1996; Greenwood, Tapia, Abbott, & Walton, 2003; Hintze, Ryan, & Stoner, 2003; Lembke, Deno, & Hall, 2003; Luze et al., 2001).

Extending CBM to Students with Significant Disabilities

The instruction for students with significant cognitive disabilities has a history focusing on functional skills, assessed primarily using mastery monitoring and task analysis. With the increased emphasis on accountability and tying instruction for students with significant cognitive disabilities to general education standards, there is a need to develop assessments that are reliable and valid but also sufficiently sensitive to the improvement of these students in academic areas, such as reading. Because of the recent shift in instruction for students with significant cognitive disabilities, with more emphasis on academics, research in the area of assessment in academic areas for these students is sparse. Al Otaiba and Hosp (2004), Tindal, McDonald, Tedesco, Glasgow, Almond, Crawford, and Hollenbeck (2003), and Wallace and Ticha (2007a) have begun work on developing progress measures in reading with these students.

Al Otaiba and Hosp (2004) used CBM as one of their assessment measures to monitor progress while implementing a tutoring model in reading with four students with Down syndrome, ages 7 to 12 years. The researchers used a pre- and post- standardized measure, the Woodcock Reading Mastery Test-Revised (WRMT-R), as well as weekly CBM to assess student progress. The materials for CBM were chosen based on the student's reading ability as measured by the Peabody Picture Vocabulary Test – Revised (PPVT-R), the Comprehensive Test of Phonological Processing (CTOPP), and the Woodcock Reading Mastery Test – Revised (WRMT-R) at the beginning of the study. The CBM materials included letter-sounds naming, sight-word identification, and oral reading fluency. Each CBM probe was administered for 1 minute. Al Otaiba and Hosp concluded that CBM in all its forms used in this study, i.e. letter-sound naming, sight-word identification and oral reading fluency, is an appropriate assessment tool for measuring progress in reading for students with Down syndrome.

Tindal et al. (2003) developed a set of standardized tasks for students with disabilities which would assess the same construct as that of large scale assessments. Even though Tindal et al. referred to these standardized tasks as CBM, the measures were designed to use general constructs rather than material directly taken from the curriculum. Four hundred and thirty seven students from Kindergarten through grade 12 were selected for the study. The selection criteria included: exemption from taking the state test, diagnosis of moderate to severe disability, and participation in a functional daily living skills curriculum. Out of the 437 students, 75 with the lowest abilities could not be assessed. In reading, students were assessed with Naming Pictures, Naming or Pointing to Letters, Blending Sounds, Reading Words, Reading Names, Reading Sentences, Reading Text Orally, and Retelling Stories as Reading or Listening Comprehension. Teachers administered only those measures that were appropriate for the student's reading level. Tindal et al. found that the CBM measures implemented in the study demonstrated sound inter-scorer reliability and construct validity. Tindal et al. concluded that the CBM measures used in there are suitable to compliment large-scale assessments to monitor student progress.

Wallace and Ticha (2007a) have conducted a pilot study on the development and validation of GOMs in reading and math for use with students with significant cognitive disabilities. Students with significant cognitive disabilities are typically those who take alternate assessments based on alternate achievement standards. The participants in the study were 13 students with significant cognitive disabilities from Kindergarten through grade 5. The newly created GOMs in this pilot study were timed identification and matching measures using pictures, letters, and sight words as stimuli. Student performance was measured at 3, 5, 7, and 10 minute time frames. Despite the pilot nature of this study with a small sample, Wallace and Ticha (2007a) found that participants were able to respond to the timed GOMs. The format and administration procedures were appropriate. Preliminary analyses suggested that Word Matching, Letter Identification, and Word Identification GOMs at 3 minutes of administration demonstrated the most promising criterion validity with the Peabody Picture Vocabulary Test –

III (PPVT-III) and the Early Literacy Knowledge and Reading Readiness Checklist (r = .46 - .92). An obvious limitation of the data produced in this pilot study was the ceiling effect on the scores resulting from a small number of test cards (27) in all of the GOMs, in combination with too much time allocated for the administration of these measures (7 and 10 minutes).

In summary of the existing studies on timed measures of reading for students with significant cognitive disabilities, GOMs showed real potential to assess reading performance for students with significant cognitive disabilities. The population of students with significant cognitive disabilities embodies a great diversity in reading ability. Further research is warranted into which GOMs and at what time frames are most appropriate for students with certain aspects of significant cognitive disabilities.

Students with Significant Cognitive Disabilities Present Unique Challenges

Students with significant cognitive disabilities typically struggle with various needs, such as the lack of verbal ability, or physical and cognitive challenges. It is therefore important to develop measures that have universal access (Gramm, 2007; Quenemoen et al., 2004). Consideration must be given to the stimulus and response requirements of this student population (Buekelman & Mirenda, 1998; Browder & Spooner, 2006; Kopperhaver, 2000; Kovach & Kenyon, 2003; and McEwen, 1997). In addition, students with significant cognitive disabilities do not typically follow a "grade-level" curriculum and little is known about their reading ability. Al Otaiba and Fuchs (2002) suggest that there is little information about predictors of future reading success for this population of students, yet we know there are significant individual differences in how students respond to instruction. Reading instruction for students with significant cognitive disabilities has most typically focused on sight words for functional reading. However, Browder, Courtade-Little, Wakerman, and Rickelman (2006) conclude that it is not necessary to choose between a functional and a literacy-based approach to reading. They suggest both can benefit students with significant cognitive disabilities in addition to information about literacy concepts. Both types of instruction should occur and technically adequate assessments should exist to measure performance and progress.

Chall's model of reading development (1996) provides a framework for recognizing the possible usefulness of measures found to be applicable for typically developing emergent and early readers as well as for students with significant disabilities in a similar stage. For example, a GOM that consists of identification of lower and upper case letters, is associated with Chall's Stage 0 (Pre-Reading) and is often used in Kindergarten may be utilized to assess the performance of a 4th grade student with significant cognitive disabilities who is at that developmental reading stage. This GOM is also aligned with one of the National Reading Panel's areas of reading instruction – phonics (see Table 1). Table 1 illustrates the connection between Chall's stages of reading development, National Panel areas of reading instruction, and the GOMs developed for this study.

The purpose of this study is to develop reliable and valid measures for students with significant cognitive disabilities that are sensitive to progress in reading. The research questions addressed in this study included:

1. Do the GOMs in reading produce reliable data when used with students with significant cognitive disabilities?
2. Do the GOMs in reading produce valid data when used with students with significant cognitive disabilities?
3. Do the GOMs in reading show growth over time when used with students with significant cognitive disabilities?

Method

Participants

The participants in the study were 31 students with significant cognitive disabilities from five urban schools in Minnesota. Twenty-two students (71%) were male and nine (29%) were female. For the purposes of this study, students with "significant cognitive disabilities" were defined as "students who participate in alternate assessment with alternate achievement standards linked to state grade level content standards" (NCLB, 2005). The 31 students were included in the study based on the following process. First, five schools with a program for students with developmental cognitive delay were identified by a teacher on special assignment in the school district. Eight teachers who teach students with significant cognitive disabilities agreed to participate in the study. Students in the eight classrooms whose parents gave permission participated in the study. Students from Kindergarten through grade twelve were represented (see Table 2 for the distributions of student grades). There were 12 (38.7 %) African American, six (19.4%) Hispanic, 12 (38.7%) White, and one (3.2%) Native American students. Twenty-two of the 31 students (71.0%) received free or reduced lunch. Five students (16.1%) were English Language Learners (ELL). Based on students' information in their IEPs, the primary disability of the students was as follows: DCD (developmental cognitive disability) was a primary label of 20 students (64.5%), two students were labeled SMI (severe multiple impairment, 6.5%), one OHI (other health impairment, 3.2%), one TBI (traumatic brain injury, 3.2%), six students had a label specific to the district, SNAP (student needing alternative program, 19.4%), and one student (3.2%) was classified as having a developmental delay in early childhood special education. In the case of the students whose primary disability label was not DCD, their secondary or tertiary label suggested this impairment.

In comparison, the demographic composition of students in special education in the school district from which the study sample was obtained was as follows: 67% male and 33% female; 53% African American, 12% Hispanic, 24% White, and 6% Native American; 73% received free or reduced lunch; and 15% were English Language Learners (ELL). Our sample was therefore a good representation of the district demographics in the categories of gender, free and reduced lunch, and ELL. In the category of ethnicity, in our sample African-American and Native American students were under-represented, while Hispanic and While students were over-represented.

Materials

The assessment materials in reading used in this study consisted of eight general outcome measures (GOMs; see Table 3) and four criterion measures. The choice of measures was based on the results of a pilot study conducted prior to this study (Wallace, & Ticha, 2007a). There was one GOM matching measure, four GOM identification measures, one GOM comprehension measure, and two GOM decoding measures. Each GOM was administered using a set of 60 laminated 8.5 x 11 inch cards that were numbered from 1 to 60. Card number 58 was a model card, cards with numbers 59 and 60 were practice cards, and cards with numbers 1 – 57 were test cards. For examples of four GOMs (Functional pictures and signs, K-4 science content word identification, Rimes, and Simple maze) see Figures 1 - 4 in Appendix 2. Each measure was accompanied by a sheet with detailed administration directions. A scoring sheet was used to record responses every time a GOM was administered. The primary and secondary data collectors used a small portable tape recorder with an ear piece and a tape with 1, 3, and 5 minute recorded time markers, along with a timer.

The four criterion measures used for this pilot study were the Peabody Picture Vocabulary Test – Third Edition (PPVT-III; Williams & Wang, 1997), the Letter-Word Identification subtest of the Woodcock-Johnson Tests of Achievement-III (WJ-LWI; Woodcock, McGrew, & Mather, 2001) the Research Institute on Progress Monitoring (RIPM) Early Literacy Knowledge and Reading Readiness Checklist – Version II (the Checklist; Wallace & Ticha, 2007b) and the Minnesota Test of Academic Achievement in reading (MTAS; Minnesota Department of Education, 2007). The PPVT-III is an untimed and individually administered assessment tool of receptive vocabulary. Assessed persons are asked to point to one black and white drawing on a page from a choice of four. It is designed for use with participants between the ages two and 90+ years. It has two parallel forms, A and B. The PPVT-III was developed in 1996 using 2,725 participants nationwide between ages two and 90+ years. Eighteen percent of the sample was African American students, 64% White, 13% Hispanic and five percent were of other origin. The participants receiving special education services in the standardization sample were: 5.5 % students with learning disabilities, 2.3% students with speech impairment, 2.2% adults with mental retardation, and 1.2% students with mental retardation. The reliability coefficients reported were above .90. To establish criterion validity of the PPVT-III, the authors used three intelligence tests: the Wechsler Intelligence Scale for Children – Third Edition (corrected correlation coefficients .91 and .92 for the two parallel forms), the Kaufman Adolescent and Adult Intelligence Test (.87 and .91), and the Kaufman Brief Intelligence Test (.82 and .80). The validity of PPVT-III was also examined using the Oral and Written Language Scales, namely Listening Comprehension (.68 and .70) and Oral Expression (.75 and .73).

The second criterion measure used was the Letter-Word Identification subtest of the Woodcock-Johnson III Tests of Achievement (WJ-LWI). The WJ-III is an assessment designed to measures academic achievement in reading, oral language, math, written language, and academic knowledge skills. The test can be used with people between 2 and 90 years. It is an un-timed test. The scores most typically used are standard scores. Normative data was collected from 8,818 participants. The sample was stratified to be representative of the U.S. population according to the 2000 census. In the Letter-Word Identification subtest**,** students are required to point and respond orally. Students are initially asked to identify letters from print, followed by words. The reported test-retest reliability for the Letter-Word Identification subtest is .95. The validity coefficient for the WJ-III Reading composite scores (broad reading and beginning reading) with reading composite scores from the Kaufman Test of Educational Achievement and Wechsler Individual Achievement Test from ranged from .44 to .82. In a school aged sample (6-19), Letter-Word Identification correlates with other WJ-III subtests from .16 to .70.

The third criterion measure used in this study was the RIPM Early Literacy Knowledge and Reading Readiness Checklist – Version II (the Checklist) for special education teachers developed originally for a pilot study preceding this study (Wallace & Ticha, 2007a). Version II of the Checklist was used for this study with small modifications based on the feedback form the teachers in the pilot study. Teachers filled in the Checklist for each student in the study. For details on the development of the Checklist see Wallace and Ticha (2007a). The Checklist consists of six subscales: I.) Concepts about books, print, letters and words, II.) Alphabetic knowledge and beginning decoding skills, III.) Phonemic awareness, IV.) Sight word vocabulary, V.) Beginning comprehension skills, and VI.) Daily living reading skills. Teachers answer each item with either a “yes” or “no” response. The number of “yes” and “no” responses is recorded for each subscale as well as total scores.

The forth criterion measure, the Minnesota Test of Academic Skills (MTAS) in reading, is an alternate assessment for students with the most significant cognitive disabilities. The MTAS is part of the statewide assessment program based on alternate achievement standards aligned with Minnesota grade-level content standards. The MTAS in reading was administered for the first time in the spring of 2007. The test is individually administered by a special educator trained to do so, and is untimed. The reading portion of the test is administered in grades 3-8 and 10 (Lombard, 2007). During the first administration, the reading portion of the test contained 12 items. The purpose of the MTAS is to sample student knowledge without having to assess on every standard or benchmark. In the future, each test will be shortened to 7 items. The test is scored using a scoring rubric that reflects the independence and correctness of the student's response along with the mode of their response. The student receives a score between 0 and 3, where 3 stands for a correct answer with no assistance except refocusing or restating and 0 stands for no attempt or response. The student is also awarded an access point A, B or C, depending on whether they respond verbally or through another mode of communication, such as pointing (Minnesota Department of Education, 2007).

Procedures

General outcome measures (GOMs) development. The GOMs were developed based on the principles of Curriculum-Based Measurement (CBM) while taking into account the skills of students with significant cognitive disabilities. The premise the GOM measures for students with significant disabilities in this study is supported in Browder et al. (2005) regarding the need to examine strategies for assessing academic progress of students with significant cognitive disabilities. For this study, the GOM measures created for the initial pilot study (Wallace & Ticha, 2007a) were modified and new ones were developed based on the same principles. The ideas for the development of GOM measures for students with significant disabilities stemmed from several sources. CBMs for typically developing students were considered as well as progress monitoring measures developed for students in early childhood education (e.g. IGDIs and DIBELS). Chall’s (1996) stages of reading development served as an organizer for the new GOMs into a sequence of reading development of typically developing students. Alternate achievement standards and alternate assessments in reading in Minnesota and other states (e.g. Massachusetts) were examined and so were progress monitoring measures for students with significant cognitive disabilities in other states, such as Oregon. In addition, several curricula and related materials for students with significant cognitive disabilities were examined, e.g. Edmark or sight word lists (Fry’s 300 Instant Sight Word List). An advisory committee meeting was held with special education teachers, specialists, researchers, and administrators to discuss the context and possibilities for developing GOMs in reading for students with significant cognitive disabilities. Conversations with special educators in the district in which this study was conducted also helped give the GOM measures their content and form.

Based on existing research and literature and also data collection, results and feedback from special educators involved in the pilot study (Wallace & Ticha, 2007a), eight GOMs were created for the purposes of the current study with the aim to assess different aspects of reading at difference levels: 1.) K-4 science content word matching, 2.) Functional pictures and signs identification, 3.) Mixed letter identification, 4.) Fry’s 100 sight words identification, 5.) K-4 science content word identification, 6.) Simple sentence maze, 7.) Rimes, and 8.) Onsets. The number of cards was increased from 30 in the pilot study to 60 and time was shortened from 10 minutes to 5 minutes in the fall and even further to 3 minutes in the spring.

The K-4 science content vocabulary matching and identification measures were created with a goal of assessing students' reading performance and progress using vocabulary words aligned with subject-specific content taught in general education classrooms across multiple grade levels. We decided to focus on the content area of science. Using the state of Minnesota's academic content standards as our framework, we identified two science strands from those standards that included science content taught across multiple grade levels. From the Earth and Space Science strand, we selected approximately 30 science content vocabulary words related to the Water Cycle, Weather, and Climate sub-strand and taught across grade levels K - 4. From the Physical Science strand, we selected approximately science content vocabulary words related to the Structure of Matter sub-strand and taught across grade levels 1 – 4.

The goal the Functional pictures and signs was to create a GOM with which to assess students' performance and progress in identifying common objects and survival signs typically found in their environment and necessary for functioning as independently as possible at home, school, in the community, and at work. Several Internet searches were conducted as we decided which categories of objects and signs to include. Eventually, we created line drawings of 39 common items used for: personal care, clothing, daily living, school, and recreation. We also created line drawings of 21 international survival signs that included outdoor survival signs typically found in community settings and indoor survival signs typically located in employment settings. When creating our line drawings of objects and survival signs, we followed Heidemarie's guidelines (1985) and focused on creating drawings with realistic details and proportions, uncluttered backgrounds, and clearly defined outlines and contours.

Based on our initial results, the Upper-case Letter Identification measure from the pilot study was changed into a Mixed-letter Identification measure by mixing lower with upper-case letters. The combinations of three upper-case letters in the letter measures were again created randomly with checks for repetitions of the same letters, letters that were visually too similar, and letter of the same case.

The Fry's 100 sight words measure from the pilot study was kept the same. The extra 30 cards were again created using a random generator with checks for repetitions and visually similar words. See Wallace and Ticha (2007a) for more detail.

In an effort to expand the components of reading development assessed through the administration of the GOM measures, we developed Rimes, Onsets and Simple Maze tasks. We followed the same protocol as we did with the development of measures in the Wallace and Ticha (2007a) pilot study. For Rimes and Onsets, we provided two words along the top of the card with a blank. The students selected by pointing the word that started (Onsets) or ended (Rimes) with the same sound as the words presented from three options. We used three or four letter words and watched for similarity or duplication in the stimulus words selected. We were careful not to use combinations of consonants that represented different blends, such as: "st" and "str." The Simple Maze is an attempt to create a measure that assesses comprehension. It requires the student to choose a word from three options that completes a 3 or 4 word sentence presented at the top of the card. Attempts were made to ensure only one word made logical sense and that the response options were not similar in sound or meaning.

Measure administration and data collection. The GOMs were individually administered by a primary data collector at a desk, in a quiet area of a school if possible. Data collectors were graduate students in education or educational psychology. All reading GOMs were administered twice in the course of the school year, in the fall and in the spring. Each student in the study was given only half or one set of the GOMs. Set 1 contained the following GOMs: Mixed letter identification, K-4 science content word matching, Onsets, and Simple sentence maze. Set 2 contained Fry's 100 sight word identification, Functional pictures and signs, Rimes, and K-4 science content word identification. The sets were created by matching the GOMs on a difficulty level and assigning each pair to a set. Students in the study were matched on grade level and pairs of students were randomly assigned either to Set 1 or 2, with one student added in the spring into Set 1. All students stayed assigned to the same set of measures both in the fall and in the spring.

During measure administration, the primary data collector placed cards in front of the student one at a time, starting with the model card. Following standardized administration directions for each GOM, the data collector modeled the task. Next, the data collector made sure the student was attentive to the task and was able to point to an item on the card. The student was asked to point to an item on the card according to the administration directions for each GOM. In the case of the two practice cards, if the student pointed to a correct item within 5 seconds as measured with a timer, the data collector administered the second practice card. If the student pointed to an incorrect item or did not point to any item, the data collector followed the prompting system described below and in the standard directions until the student pointed to the correct practice item. If the student pointed to the correct item on the second practice card within 5 seconds as measured with a timer, the data collector administered the first test item, while starting the tape recorder functioning as a timer (a tape with recorded 1-minute [in the spring only], 3-minute, and 5-minute [in the fall only] markers). As the student responded to the two practice cards and the set of 57 test cards, the data collector recorded the responses, including a prompting level required, on a scoring sheet. In addition to recording a correct or incorrect response and a corresponding prompting level, the primary data collector recorded any behavior directives they used if a student strayed off task. The primary data collector made a mark on the scoring sheet at 1 minute (spring only), 3 minutes, and 5 minutes (fall only). All the cards for all GOMs were always administered in the same order.

Most of the time, a secondary data collector was present to shadow the first data collector in recording the student's responses, prompting levels, behavior directives, time intervals, and other observations in order to check for accuracy, to record additional information on the student's testing behavior, and to generally assist the primary data collector. The same GOMs with cards in the same order were administered in fall and in the spring. All data collectors participated in a training session prior to data collection that addressed not only the administration procedures of the measures but also contained information about working in the schools with students with significant cognitive disabilities.

The WJ-LWI and the PPVT-III were individually administered according to the standardized directions of the tests in the same setting as the GOMs by the primary data collector. The two criterion measures were not timed and were administered to all students in the study in the fall and spring. Two different forms of the WJ-LWI and the PPVT-III were administered in the fall and spring. The special education teachers in the study were given the Checklist to complete for all their students in the fall and in the spring. Additional data on the students in the study was collected from the district data base in the form of demographic data, such as grade, SES and ELL status, as well as IEP goals and objectives in reading. The Minnesota Test of Academic Skills (MTAS) raw scores were obtained from the district.

Scoring sheet. Student responses on the GOMs, including basic demographic information, were recorded on scoring sheets. Every GOM administration required a separate scoring sheet. The scoring sheet was common across all the GOM measures. The top of the front page of the scoring sheet contained a list of GOMs to check for the one administered, a space for information about the student, the date of measure administration, and the name of the data collector administering the measure, as well as the observer. Below, there was a space for recording a student's primary mode of pointing. Then on the front page of the scoring sheet, there was a line for each of the 60 items administered with scoring and prompt level options to circle, space for notes and space for tallying behavioral directives. Also on the front page, the data collectors circled 0 or 1 for an incorrect or correct response on the two practice items (cards 59 and 60) along with a level of prompt the student needed to give a correct response (0, 1, 2, or 3). On the next four pages, the data collectors recorded 0 or 1 for incorrect or correct responses and 0, 1, 2, or 3 for the level of prompt used for each test item in case the student did not respond to a card, any observations of the student testing behavior and behavior directives, if used. The total number of test cards was 57.

Prompting system. In order to ensure that all students in the study were able to respond to the items on the GOMs, a four-level prompting system developed for a pilot study preceding this study (Wallace & Ticha, 2007a) was used. Level 0 prompts were non-prompted, Level 1 prompts were verbal with directions repeated once, Level 2 prompts had verbal and gestural (pointing to the correct item) components, and Level 3 prompts, provided when students were not able to respond to practice items, included both verbal and partial physical (guiding the hand of the student to point to the correct item) components. See Wallace and Ticha (2007a) for a detailed description of the prompting system.

Scoring. The GOMs were scored by counting and recording the number of correct (1) and incorrect (0) responses for each time frame on the test part of the scoring sheet. The number of each level of prompt (0-3) used was also recorded. In the analyses, only those responses that did not require a prompt were used. In addition, all GOM scores were corrected for guessing using a 3-consecutive-error rule. According to this rule, only those scores were counted as correct that preceded three consecutive incorrect responses. This particular scoring rule was implemented based on a standard practice used with CBM maze selection measures as well as on the results of an empirical comparison of three scoring rules, the 3-error rule, 5-error rule and formula scoring (Mehrens & Lehman, 1991) conducted using the data from this same study. One of the criterion measures, the RIPM Checklist, was scored by counting the number of "yes" and "no" responses for each of the six subscales and in total. The WJ-LWI and PPVT-III criterion measures were scored according to the standardized published directions. Finally, the number of correct responses using only minimal level of assistance, i.e. "Correct Response: No assistance provided other than refocusing or restating" (Minnesota Test of Academic Skills [MTAS] Scoring Rubric), or score of 3 was used when analyzing the MTAS data.

Analysis. The number of correct responses not requiring a prompt and corrected for guessing was used as the unit of analysis for the GOMs. The number of "yes" responses was used as a unit analysis for the Checklist. Standard scores were used for analysis for both the WJ-LWI and the PPVT-III. The data was analyzed using descriptive statistics as well as inferential statistics. Means and standard deviations for all time frames of the GOMs, the Checklist, the WJ-LWI, the PPVT-III, and the MTAS were computed. GOM scores recorded at 3 minutes were used for inferential statistics because that was the common time frame in the fall and spring. Inter-observer reliability was calculated for the GOMs. Inter-scorer and test-retest reliabilities were calculated for the GOMs and criterion measures. Pearson correlations were computed between the GOMs and all criterion measures in the fall and spring in order to establish criterion validities for the GOMs. The differences between the mean fall and spring scores on the GOMs and the criterion measures, except the MTAS, were calculated using a series of t-tests to find out whether the measures and the GOMs in particular were sensitive to growth over time for students with significant cognitive disabilities.

Results

This study addressed three research questions:

1. Do the GOMs in reading produce reliable data when used with students with significant cognitive disabilities?
2. Do the GOMs in reading produce valid data when used with students with significant cognitive disabilities?
3. Do the GOMs in reading show growth over time when used with students with significant cognitive disabilities?

Descriptive statistics for all the measures used in the study are presented in Tables 4 (fall) and 5 (spring). Prior to calculating both descriptive and inferential statistics, all GOM scores were adjusted for prompting and guessing, by removing all the scores that required a prompt higher than 0 and by using a rule of stopping counting the number of correct responses after the student made three consecutive errors respectively. In the fall, the GOMs were administered for 5 minutes and a score was also recorded at 3 minutes. Based on the data from the fall, the spring GOMs were administered only for 3 minutes with a score recorded at 1 minute. The same version of GOM measures was administered in the fall and spring. Form A of the WJ-LWI was administered in the fall, while Form B was administered in the spring. In the case of the PPVT-III, Form B was administered in the fall while Form A was administered in the spring. Raw scores were used for analyzing all measures except the WJ-LWI and PPVT-III, in which case standard scores were used.

The numbers in bold in the Minimum and Maximum columns in Tables 4 and 5 indicate floor and ceiling effects. When the time of administration was shortened from 5 to 3 minutes, only one GOM (Fry's 100 sight word identification) showed a ceiling effect in contrast to three GOMs in the fall. In addition, the shortened administration time made a difference in how many measures showed a floor effect, from only three GOMs in fall to five in the spring.

The distributions of the measures were in the normal range. For the measures administered in the fall, skewness ranged from -1.00 (Functional Pictures and Signs) to .64 (Rimes) for GOMs, and -.63 (WJ-III Letter-Word Identification) to .97 (PPVT-III) for criterion measures. Kurtosis ranged from -1.93 (K-4 Science Content Word Matching) to .56 (Functional Pictures and Signs) for GOMs, and from -1.31 (the Checklist) to .17 (PPVT-III) for criterion measures. Similarly in the spring, skewness ranged from -.55 (Functional Pictures and Signs) to .75 (Onsets) for GOMs, and -.70 (WJ-III Letter-Word Identification) to .60 (PPVT-III) for criterion measures. Kurtosis in the spring ranged from -1.67 (Rimes) to -.10 (Functional Pictures and Signs) for GOMs, and from -1.09 (MTAS) to -.22 (WJ-III Letter-Word Identification) for criterion measures.

In the case of all the GOMs in both time frames, there was an increase of correct responses between 3 minutes and 5 minutes and 1 minute and 3 minutes, suggesting that the measures can be used with students with significant cognitive disabilities to detect timed progress.

Reliability

Three types of reliability were calculated on the data. First, inter-observer reliability between the primary and secondary data collectors was calculated separately for fall and spring data collections. Only scores for true reliability were included, in that inter-observer reliability was only considered true when the primary and secondary data collectors did not check or consult their scoring sheets during or after data collection. True reliabilities were calculated on approximately 25% of fall GOMs and on 43% of spring GOMs. The true inter-observer reliability was 99.63% for fall and 99.38% for spring.

Second, inter-scorer reliability was calculated on approximately 27% of fall and 25% of Spring GOMs. The inter-scorer reliability was 99.50% in the fall and 96.63% in the spring. Inter-scorer reliability was also calculated on 20% and 19% of the three criterion measures, the WJ-LWI, PPVT-III, and the Checklist in the fall and spring respectively. The inter-scorer reliability for fall and spring was 100% for all criterion measures.

Third, in order to find out whether the students in the study performed consistently on the same measures at the two different time frames, test-retest reliability was calculated on all the measures. The scores at 3 minutes were used for reliability calculations on the GOMs because that was the only time frame common to fall and spring administrations. The test-retest reliability coefficients for GOMs ranged from .75 (Rimes) to .94 (K-4 Science Content Word Identification). The test-retest reliability for the criterion measures ranged from .84 (PPVT-III) to .95 (the Checklist). All correlations were significant at the .01 level.

Validity

In order for special education teachers to use the newly developed GOMs with confidence, first it needs to be established that the GOMs measure concepts that are relevant; and second that the GOMs measure concepts that are within the range of ability for students with significant cognitive disabilities. To help establish the relevance and suitability of the GOMs for students with significant cognitive disabilities, two types of validity were calculated on these measures - concurrent and predictive. The results of concurrent validities in the fall and in spring are presented in Tables 6 and 7. The GOM scores at 3 minutes were used for the analysis. The sample size column (*n*) in Table 7 contains two numbers because not all students that were administered the GOMs were given the MTAS. The administration of the MTAS is based on state requirements for certain grades. Based on the concurrent validity results in the fall (Table 6) and in the spring (Table 7), four GOMs appear to have the best potential for use as efficient indicators of student performance in the classroom, Rimes, K-4 Science Content Word Identification, Simple Maze, and Fry's 100 sight word identification. Even though Simple Maze did not relate as well to the PPVT-III or even the WJ-LWI subtest in the spring, the fact that it related well to the more challenging MTAS can give confidence for its potential use with older or more advanced students. Although the Functional Pictures and Signs related well only to the PPVT-III, it suggests that this GOM is probably more suited for use with younger or less advanced students with significant cognitive disabilities. The one matching measure, K-4 Science Content Word Matching showed the least promise. Onsets as well as Mixed Letter Identification, although not performing as well as the other GOMs across all the criterion measures, related well to the Checklist both in the fall and spring. Apart from the Functional Pictures and Signs, Rimes and both word identification measures related well to the PPVT-III. As expected, the two word identification measures related well to the WJ-LWI. In general, the GOMs related the best to the Checklist completed by the special education teachers. That means that student performance on the GOMs reflects teachers' evaluation of their students' reading progress. Because special education teachers are with their students most of the time and their class size is small, their judgment adds value to the GOMs. The results with MTAS as a criterion measure need to be interpreted with caution because the MTAS scores used in the study were the scores of its first district-wide administration and also because only the raw scores of correct responses with minimal assistance were used for analysis.

Predictive validity based on student scores on GOMs in the fall and criterion measures in the spring is presented in Table 8. The pattern of results for predictive validity follows the pattern of results for concurrent validity in Tables 6 and 7. If a teacher was to administer GOMs in the fall to get an idea how the student will perform on the criterion measures in the spring, he or she should probably choose Rimes, K-4 Science Content Word Identification, or Fry's 100 sight word identification if he or she wanted to test the student on the PPVT-III, the WJ-LWI and complete the Checklist in the spring. The teacher can choose the Functional Picture and Signs measure if he or she just wants to know how the student will do on the spring PPVT-III. Simple maze would give the teacher an idea how the student would perform on the state assessment in reading (MTAS). Only the Onsets GOM did not follow the pattern of concurrent validity. Table 8 shows that along with the K-4 Science Content Word Matching, Onsets is the weakest GOM across all the criterion measures.

Growth

The growth in reading scores of students with significant disabilities on all GOMs and three criterion measures, PPVT-III, WJ-LWI and the Checklist, between fall and spring was examined in order to find out whether the GOM measures were able to detect growth and what growth they detected in relation to the criterion measures. Figure 5 offers a visual representation of reading growth on all measures. The GOM scores plotted in Figure 3 are scores recorded at 3 minutes and adjusted for prompting as well as for guessing using the 3-cosecutive-error rule.

Based on the results of a series of t-tests, growth on all but two GOMs, Rimes and Onsets, was significant at $p < .05$ level. Growth on four measures, Functional Pictures and Signs, K-4 Science Content Word Identification, Mixed Letter Identification and Fry's 100 sight word identification, was significant at $p < .01$ level. One criterion measure, the Checklist, showed significant growth over time ($p < .01$).

The rate of growth on Functional Pictures and Signs, Mixed Letter Identification and Fry's 100 sight word identification - the GOMs that showed the most significant growth - was between 1.2 and 1.4 pictures, letters or words a month. In comparison with the recommended rates of growth on CBMs for typically developing students, the rates of growth found for students with significant disabilities are slower. More research is needed to replicate these findings.

In addition to showing the rate of growth, Figure 5 provides a visual distinction between student mean performance on the decoding and comprehension GOMs (Rimes, Onsets and Simple Maze) and the rest of the GOMs, suggesting that the former are more difficult than the latter.

Discussion

The importance of assessing reading performance and progress for all students has become intensified as the stakes become increasingly higher. All students must be included in accountability systems and all students must meet standards that align with grade level standards in reading, but how? Assessing the performance of students with significant cognitive disabilities has long been a challenge, especially in academic content areas, such as reading.

Students, who are often nonverbal, present significant cognitive and sometimes physical challenges, can find standardized tests, paper and pencil tests, and verbal response requirements impossible. In addition, clarity about what and how they should learn has not been determined. However, they must be included in the accountability systems in a way that is respectful, relevant, reliable and related to alternate achievement standards linked to grade level standards.

Clearly needs exist in this area; however, assessment and expected progress in academic content for students with significant cognitive disabilities are areas that have largely remained unexamined. Recent federal and state requirements to ensure all students are progressing and meeting state standards draws attention to students with such challenges. While portfolios and mastery monitoring strategies, (e.g. a checklist), have been used as alternate assessments in some states, e.g. Massachusetts and Nebraska respectively, each experiences its challenges. In the case of portfolios, two of the biggest drawbacks are the time spent creating a portfolio and using a portfolio for measuring student progress. The main challenge of using a mastery monitoring approach is again measuring student progress that goes beyond a single skill. The goal of this study was to develop general outcome measures for students with significant cognitive disabilities that were time efficient, were reliable and valid, and had the potential of measuring student progress across time. Using Chall's model of reading development and previous experience with curriculum-based measurement (CBM), it was anticipated that general outcome measures (GOMs) could be developed and used with students with significant cognitive disabilities. A pilot study provided empirical support for using these newly created GOM measures (Wallace & Ticha, 2007a). The present research was conducted to determine if using newly developed general outcome measurement could potentially work to assess students' performance in reading or early literacy.

Chall (1996) provides a framework for thinking about reading development that allows for such development to be decoupled from age, which suggests that even an older student may be at a very early stage of reading development. This perspective supports the approach we have examined in our research – that general outcome measures (GOMs) can be used to measure *early* literacy development with *older* students who have significant cognitive disabilities. The hypothesis was that such measures could be created to measure students' performance in academic areas aligned with state standards as well as IEP goals; ultimately, providing teachers with a tool to measure individual annual growth. The research questions posed for this study focused on three areas: The first set of research questions addressed whether the newly developed GOM measures produced reliable data when used with students with significant cognitive disabilities, while the second set of questions examined the concurrent and predictive validity of the data produced using these measures. Finally, the new measures were studied to determine if they would show growth over time when used with students with significant disabilities. The results suggest GOMs may be an appropriate way to measure the performance and growth of students with significant cognitive disabilities in an academic area, such as reading.

The administration of the measures using laminated cards that required only a "pointing" response worked well for students to engage in the task. While we started with 5-minute measurement intervals, it was clear that students could generate responses within 3 and even 1 minute. Therefore, analyses were conducted using the 3-minute samples. Up to a point, the more time given to complete the task, the greater number of correct responses students gave, which suggests timed measures can be used with students with significant cognitive disabilities. One of the criteria for such measures as noted by Deno (1985) is practicality of use by teachers. The data at 3 minutes support the use of timed measures with students with significant cognitive disabilities; which is contrary to past belief, but gives increased evidence for the practicality of the measures for use with students facing similar challenges.

The students required demonstration, training and practice in order to appropriately respond to the measures; however, very little formal prompting was needed with this sample of students as was the case in the prior pilot study (Wallace & Ticha, 2007a). The scores used in analysis were corrected for prompting and guessing as described earlier in the article. In addition to establishing a level of practicality for teachers, measures must also be reliable and valid as noted by Deno (1985).

Reliability results from this study suggest initial support for the use of GOM measures with students with significant cognitive disabilities. The results gave a positive indication that a person administrating the GOMs to a student in a school setting can have confidence that scores he or she records based on student responses will be reliable representations of the student’s performance on that particular measure. In addition, the inter-scorer reliability suggests that a student’s responses to the GOMs can be scored with high level of accuracy by the scorer. The test-retest reliability coefficient range for GOMS from .75 (Rimes) to .94 (K-4 Science Content Word Identification) suggests that a special education teacher can have confidence in predicting his/her students’ performance on a GOM administered in the spring based on the students’ performance on the same GOM in the fall. These findings suggest that the GOMs reliably measure performance of the students.

To establish the validity of the GOMs for students with significant cognitive disabilities, two types of validity were examined, concurrent and predictive. It is important to keep in mind that in order to establish validity of a new measure, the researcher needs to have some confidence that the criterion measures represent the concepts or the aspects of a concept that the researcher is trying to assess. The criterion measures in reading or pre-reading skills designed specifically for students with significant cognitive disabilities are very limited. That was the reason the Checklist to be completed by special education teachers was developed for the purposes of this study. The picture criterion, the PPVT-III, the letter-word criterion (WJ-LWI), the alternate assessment (MTAS), and Checklist all measure related but different aspects of reading skills. It was a goal for all the criterion measures as well as for the GOMs in the study to add to a construct of which reading skills should be included in the instruction for students with significant cognitive disabilities. It is clear that that there are gaps in that construct as measured by the norm-referenced criterion measures in the study. The Checklist, although with limited data on its technical adequacy, provides the most complete representation of the construct of reading skills for these students. Consequently, when interpreting the validity of the GOMs, it is necessary to keep the process in which the validity was established in mind. The GOMs that have shown the best potential for predicting the reading performance for students with significant cognitive disabilities were Rimes, K-4 Science Content Word Identification, Simple Maze, and Fry’s 100 sight word identification.

The results on the sensitivity of the GOMs to measuring reading growth over time demonstrated that growth on Functional Pictures and Signs, Mixed Letter Identification, and Fry's 100 sight word identification was greatest, with the rate of growth between 1.2 and 1.4 pictures, letters or words a month. In comparison with the recommended rates of growth on CBMs for typically developing students, the rates of growth found for students with significant disabilities are slower. Nevertheless, the results were significant at the $p < .01$ level and thus indicate that, with replication of these results, the three and all other GOMs, except Rimes and Onsets, have potential to help special education teachers assess student growth in reading. Even in the case of Rimes and Onsets, because they fall into a category of more difficult GOMs as indicated by the lower means in Figure 5, there is a possibility that when used with older or higher achieving students with significant cognitive disabilities, these two measures will also be able to detect progress.

So, what does all this mean for assessing the progress and performance of students with significant cognitive disabilities? We believe this work is good news for teachers, students, parents and others who are interested in assessment. These newly created GOMs appear to be valid, reliable measures that show growth in early literacy, an important academic area in which all students are assessed. The measures are related to reading as assessed with norm-referenced measures and reflect the opinions of the teachers who work with students closely every day, often for multiple years. While the way students are assessed within state accountability systems is through alternate assessments administered once a year, and educators and others have expressed concern regarding the assessments; the GOMs give teachers a way to measure student performance more frequently, and thus give them multiple opportunities to assess student strength and weaknesses and modify their instruction accordingly. In addition, the GOMs are designed to measure growth over time and might, in fact, not only add value to the present alternate assessment process used in many states, but also be a valuable indicator for special education teachers in the course of the whole school year, of how well each student will perform on an alternate test. Thus, we believe that the GOMs can provide teachers with an assessment tool in the classroom to help them determine if and how well students are learning. We consider the outcomes of this work to be a big step for students with significant cognitive disabilities; one that values their participation in an accountability system and provides a way to accurately assess their performance and progress in early literacy.

Limitations and Future Directions

When interpreting the results presented in this paper, it is important to do so in the context of several limitations of this study. First, the students in sample were of different grade and ability levels. The sample was small, especially because half of the students were administered half of the GOMs in the study. It is plausible that when a new study is conducted with a more homogenous sample of elementary students or secondary students, certain GOMs will perform better than others. Ideally, future studies will be able to establish which measures work better for younger or lower achieving students and which for older or higher achieving students with significant cognitive disabilities. Secondly, only one subtest of the Woodcock-Johnson III Tests of Achievement was used in this study, thus giving a limited opportunity for conclusions. The Minnesota Test of Academic Skills (MTAS) was used in its first district administration form and with a sample even smaller than in the case of the GOMs because it is not administered in all grades. Thirdly, only the GOM results at 3 minutes were used for analysis in this study because this time was common to both fall and spring data collections. One-minute administration results ought to also be explored in the future as they would be free of a ceiling effect. Despite these limitations, the results produced by this study offer another important step in progress monitoring research in reading for students with significant cognitive disabilities.

Author Note

Teri Wallace, Renáta Tichá, and Kathy Gustafson are at the Institute on Community Integration, University of Minnesota. We wish to thank DCD teachers in Minneapolis Public Schools, MN.

Address correspondence to Teri Wallace, 111A Pattee Hall, 150 Pillsbury Dr. SE, Minneapolis, MN, 55455, walla001@umn.edu.

The Research Institute on Progress Monitoring at the University of Minnesota is funded by the U.S. Department of Education, Office of Special Education Programs (Award H324H030003) and supported the completion of this work.

References

Al Otaiba, S., & Hosp, M. (2004). Providing effective literacy instruction to students with Down syndrome. *TEACHING Exceptional Children, 36*(4) , 28-35.

Allinder, R. M., & Siegel, E. (1999). "Who is Brad?" Preservice teacher's perception of summarizing assessment information about a student with moderate disabilities. *Education and Training in Mental Retardation and Developmental Disabilities, 34*(2), 157-169.

Allinder, R. M., & Swain, K.D. (1997). An exploration of the use of Curriculum-Based Measurement by elementary special educators. *Diagnostique, 23*(2), 87-104.

Browder, D. M., Courtade-Little, G., Wakerman, S., & Rickelman, R.J. (2006). From sight words to emerging literacy. In: D. M. Browder & F. Spooner (2006). *Teaching Language Arts, Math, & Science to Students with Significant Cognitive Disabilities*. Baltimore: Paul H. Brookes.

Browder, D. M. & Spooner, F. (2006). *Teaching Language Arts, Math, & Science to Students with Significant Cognitive Disabilities*. Baltimore: Paul H. Brookes.

Buekelman, D., & Mirenda, P. (1998). *Augmentative and alternative communication: Management of severe communication disorders in children and adults*. Baltimore: Paul H. Brookes.

Calhoun, M.B., & Fuchs, L.S. (2003). The effects of peer-assisted learning strategies and Curriculum-Based Measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24,* 235-245.

Chall, J.S. (1996). *Stages of reading development* (2nd Ed.). Fort Worth, Texas: Harcourt Brace.

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

Deno, S. L. (1990). Individual differences and individual difference: The essential difference of special education. *The Journal of Special Education, 24,* 160-173.

Deno, S.L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education, 37*(3), 184-192.

Duncan, D., Sbardellati, E., Maheady, L., & Sainato, D. (1981). Nondiscriminatory assessment of severely physically handicapped individuals. *Journal of The Association of the Severely Handicapped, 6*, 17-22.

Espin, C.A. & Deno, S.L. (1994-95). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20,* 121-142.

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *Journal of Special Education, 34,* 140-153.

Espin, C., Wallace, T., Cambell, H., Lembke, E.S., Long, J.D., & Ticha, R. (in press). Curriculum-Based Measurement in writing: Predicting the success of high school students on state standards tests. *Exceptional Children*.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *Journal of Special Education*, *35*(1), 4-16.

Fuchs, L. S., & Fuchs, D. (2002). Curriculum-Based Measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders, *Peabody Journal of Education, 77,* 64-84.

Fuchs & Fuchs (in press). Determining adequate yearly progress from kindergarten through grade 6 with Curriculum-Based Measurement. *Assessment for Effective Intervention*.

Good, R. H. I., & Kaminiski, R. A. (1996). Assessment of instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11,* 326-336.

Gramm, S. (2007). *Cracking the Code: IDEA and NCLB Alternate Assessment Rules Made Simple.* Horsham, PA: LRP Publications.

Greenwood, C. R., Tapia, Y., Abbott, M., & Walton, C. (2003). A building-based case study of evidence-based literacy practices: Implementation, reading behavior, and growth in reading fluency, K-4. *Journal of Special Education, 37,* 95-110.

Heidemarie, A. (1985, March). *Ways to communicate with children who are profoundly mentally retarded and cannot speak.* Paper presented at the World Congress of the International Association for the Scientific Study of Mental Deficiency, New Delhi, India.

Hintze, J.M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the comprehensive teat of phonological processing. *School Psychology Review, 32,* 541-556.

Individuals with Disabilities Education Improvement Act of 2004, P. L. No. 108-446, 20 U.S.C. section 611-614.

Kopperhaver, D. A. (2000). Literacy in AAC: What should be written on the envelope we push? *Augmentative and Alternative Communication, 16,* 270-279.

Kovach, T. M., & Kenyon, P. B. (2003). Visual issues and access to AAC. In J.C. Light, D. R. Beukelman, & J. Reichle (Eds.*), Communicative competence for individuals who use AAC: From research to effective practice* (pp. 277-319). Baltimore: Paul H. Brookes.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention*, *28*(3-4), 23-35.

Lombard, T.J. (2007). Alignment of the Minnesota Test of Academic Achievement Skills (MTAS) to the Minnesota Academic Standards. Retrieved October 2007, http://education.state.mn.us/MDE/Accountability_Programs/Assessment_and_Testing/Assessments/MTAS/MTAS_General_Information.

Luze, G.J., Linebarger, D.L., Greenwood, C.R., Carta, J.J., Walker, D., Leitchuh, C., & Atwater, J.B. (2001). Developing a general outcome measure of growth in the expressive communication of infants and toddlers. *School Psychology Review, 30*(3), 383-406.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York: Guilford.

McEwen, I. R. (1997). Seating, other positioning and motor control. In L. Lloyed, D. Fuller, & H. Arvidson (Eds.), *Augmentative and alternative communication: A handbook of principles and practices* (pp. 280-298). Needham Heights, MA: Allyn & Bacon.

Mehrens, W.A., & Lehman, I. J. (1991). *Measurement and evaluation in Education and Psychology* (4th ed.). Orlando: Harcourt Brace College Publishers.

Minnesota Test of Academic Skills (MTAS) Scoring Rubric. Retrieved June 2007, from http://education.state.mn.us/mdeprod/groups/Assessment/documents/Instruction/.

Quenemoen, R., Thurlow, M., Moen, R., Thompson, S., & Morse, A. B. (2004). *Progress monitoring in an inclusive standards-based assessment and accountability system* (Synthesis Report 53). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

No Child Left Behind Act of 2001. (2002). Pub. L. No. 107-110, 115 Stat. 1425.

Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education*, *23*(5), 300-316.

Roach, A.T., & Elliott, S. N. (2006). The influence of access to general education curriculum on alternate assessment performance of students with significant cognitive disabilities. *Educational Evaluation and Policy Analysis, 28* (2), 181-194.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for Curriculum-Based Measurement of Reading Growth. *Journal of Special Education, 34,* 164-172.

Sigafoos, J., Cole, D. A., & McQuarter, R. (1987). Current practices in the assessment of students with severe handicaps. *Journal of the Association for Persons with Severe Handicaps, 12*(4), 264-273.

Stecker, P.M., Fuchs, L.S., & Fuchs, D. (2005). Using Curriculum-Based Measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., & Crawford, L. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*, 481-494.

U.S. Department of Education. *No Child Left Behind.* http://www.ed.gov/nclb/

U.S. Department of Education (2006). The 26th Report to Congress on the Implementation of the Individuals with Disabilities Education Act. Washington, D.C.: Author.

Wallace, T., & Ticha, R. (2007a). *General Outcome Measures for Students with Significant Cognitive Disabilities: Pilot Study*. Unpublished manuscript, University of Minnesota.

Wallace, T.,& Ticha, R. (2007b). Early Literacy Knowledge and Reading Readiness Checklist – Version II. Unpublished assessment.

Wayman, M.M., Wallace, T., Wiley, H.I., Ticha, R., & Espin, C.A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*(2), 85-120.

Williams, K.T., & Wang, J. (1997). Technical References to the Peabody Picture Vocabulary Test – Third Edition (PPVT-III). Circle Pines, MN: AGS.

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

Table 1: Chall's Stages of Reading Development, Associated Ages, NRP Areas of Reading Instruction and Related Study Measures

Chall's (1996) Stages	Typical Ages Associated with Chall's Stages*	National Reading Panel - Areas of Reading Instruction	Related Study Measures
Stage 0 - Pre-reading - The learner pretends to read and relies on pictures for understanding the story context. They can name most letters, along with some words/signs from environmental print.	Birth to age 6	Phonemic Awareness	• Mixed letter ID • Functional pictures and signs • K-4 Science word ID/Matching
Stage 1 - Initial reading and decoding - The learner develops letter-sound relationships and is able to read high-frequency words by sight. The focus at this stage is on decoding rather than meaning.	Ages 6 - 7	Phonics	• Fry word ID • Rimes • Onsets
Stage 2 - Confirmation and fluency - The focus shifts from identifying single words to reading fluently.	Ages 7 - 8	Fluency	
Stage 3 - Reading to learn the new - There is a shift from learning to read to reading to learn and acquire new knowledge. The focus is self-centered.	Ages 8 - 14	Vocabulary and Text Comprehension	• Simple maze
Stage 4 - Multiple viewpoints - There is a shift from single, self-center viewpoint to multiple points of view. Thinking critically about ideas.	Ages 14 - 18	These two stages are not directly addressed in the proposed research.	
Stage 5 - Construction and reconstruction - The reader can synthesize the work of others and form his or her own opinion based on facts.	Ages 18 and older		

Table 2

Demographics by Grade

Grade	Frequency	Percent
0	1	3.2
1	2	6.5
2	2	6.5
3	5	16.1
4	6	19.4
5	4	12.9
6	1	3.2
7	4	12.9
9	1	3.2
10	2	6.5
11	1	3.2
12	2	6.5
Total	31	100

Table 3

Types of General Outcome Measures in Reading

#	Type	GOM Measures
1.	Matching (F, S)	K-4 science content words
2.	Identification (F, S)	Pictures and signs functional
3.	Identification (F, S)	Letters lower and upper case mix
4.	Identification (F, S)	Fry's 100 instant sight words
5.	Identification (F, S)	K-4 science content words
6.	Comprehension (F, S)	Simple sentence maze
7.	Decoding (F, S)	Rimes
8.	Decoding (F, S)	Onsets

Table 4

Descriptive Statistics: Reading Fall '06

Measure	Mean	SD	Minimum	Maximum	*n*
Functional pictures and signs 3 min c	24.64	10.73	1	43	14
Functional pictures and signs 5 min c	39.93	16.94	1	**57**	14
Rimes 3 min c	7.71	7.15	**0**	19	14
Rimes 5 min c	12.07	12.13	**0**	32	14
Onsets 3 min c	12.36	9.04	2	26	11
Onsets 5 min c	19.55	15.08	2	42	11
K-4 science word matching 3 min c	21.14	14.21	1	40	14
K-4 science word matching 5 min c	33.29	23.14	1	56	14
K-4 science word identification 3 min c	21.21	14.07	**0**	40	14
K-4 science word identification 5 min c	34.00	22.94	**0**	56	14
Simple maze 3 min c	10.07	7.31	2	22	14
Simple maze 5 min c	15.86	12.39	2	35	14
Mixed letter identification 3 min c	20.33	13.97	**0**	39	15
Mixed letter identification 5 min c	32.40	22.88	**0**	**57**	15
Fry's 100 word identification 3 min c	22.50	15.16	1	44	14
Fry's 100 word identification 5 min c	34.14	22.02	1	**57**	14
W-J word id SS	44.03	22.22	**0**	79	29
PPVT-III SS	51.40	12.73	**40**	86	30
Checklist total "yes"	33.37	17.13	3	56	30

Note: W-J word id SS = Woodcock-Johnson word identification standard score; PPVT-III SS = Peabody Picture Vocabulary Test standard score; c = correct; results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 6

Descriptive Statistics: Reading Spring '07

Measure	Mean	SD	Minimum	Maximum	*n*
Functional pictures and signs 1 min c	11.73	5.16	**0**	19	15
Functional pictures and signs 3 min c	32.53	14.81	**0**	51	15
Rimes 1 min c	4.36	3.10	1	9	14
Rimes 3 min c	10.36	9.74	1	26	14
Onsets 1 min c	4.00	3.16	**0**	10	15
Onsets 3 min c	10.73	10.96	**0**	31	15
K-4 science word matching 1 min c	9.27	5.33	1	19	15
K-4 science word matching 3 min c	26.80	16.73	2	54	15
K-4 science word identification 1 min c	10.14	5.96	2	19	14
K-4 science word identification 3 min c	26.64	18.35	2	50	14
Simple maze 1 min c	5.73	2.49	**0**	9	15
Simple maze 3 min c	12.40	7.75	**0**	23	15
Mixed letter identification 1 min c	10.69	6.63	**0**	20	16
Mixed letter identification 3 min c	30.81	20.59	**0**	56	16
Fry's 100 word identification 1 min c	10.87	6.59	**0**	24	15
Fry's 100 word identification 3 min c	29.00	20.62	**0**	**57**	15
W-J word id SS	46.90	21.43	**0**	79	31
PPVT-III SS	52.68	12.48	**40**	80	31
Checklist total "yes"	36.10	14.24	6	55	31
MTAS reading c with min. assist.	6.65	3.18	2	**12**	17

Note: W-J word id SS = Woodcock-Johnson word identification standard score; PPVT-III SS = Peabody Picture Vocabulary Test standard score; MTAS = Minnesota Test of Academic Skills; c = correct; results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 7

Concurrent Criterion Validity: GOM Reading Measures with PPVT-III , W-J word id, Checklist Fall' 06

GOM measure	PPVT SS	W-J word id SS	Checklist II	*n*
Functional pictures and signs 3 min c	.82**	.31	.42	14
Rimes 3 min c	.79**	.63*	.61*	14
Onsets 3 min c	.62*	.23	.77**	11
K-4 science word matching 3 min c	.25	.11	.68**	14
K-4 science word ident. 3 min c	.58*	.70**	.89**	14
Simple maze 3 min c	.60*	.55*	.86**	14
Mixed letter identification 3 min c	.30	.22	.69**	15
Fry's 100 word identification 3 min c	.61*	.58*	.83**	14

* = correlation significant at .05 level, ** = correlation significant at .01 level

PPVT SS = Peabody Picture Vocabulary Test standard score; W-J word id SS = Woodcock-Johnson letter-word identification subtest standard score; Checklist = Early Literacy Knowledge and Reading Readiness Checklist, Version II; c = correct; results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 8

Concurrent Validity: GOM Reading Measures with PPVT-III , W-J word id, Checklist and MTAS Spring '07

GOM measure	PPVT SS	W-J word id SS	Checklist II	MTAS C	*n*
Functional pictures and signs 3 min c	.71**	.44	.43	.26	15,9
Rimes 3 min c	.63*	.52	.71**	.70	14,8
Onsets 3 min c	.41	.18	.69**	.77*	15,8
K-4 science word matching 3 min c	.34	.28	.44	.35	15,8
K-4 science word identification 3 min c	.62*	.76**	.83**	.27	14,8
Simple maze 3 min c	.22	.45	.87**	.82*	15,8
Mixed letter identification 3 min c	.45	.50	.67**	.03	16,8
Fry's 100 word identification 3 min c	.66**	.76**	.82**	.47	15,9

* = correlation significant at .05 level, ** = correlation significant at .01 level

PPVT SS = Peabody Picture Vocabulary Test standard score; W-J word id SS = Woodcock-Johnson word identification subtest standard score; Checklist II = Early Literacy Knowledge and Reading Readiness Checklist, Version II, MTAS C = Minnesota Test of Academic Skills correct with minimal assist.; c = correct; results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 9

Predictive Validity: GOM Reading Measures Fall 06 with PPVT-III , W-J word id, Checklist and MTAS Spring '07

GOM measure	PPVT SS	W-J word id SS	Checklist II	MTAS C	*n*
Functional pictures and signs 3 min c	.84**	.30	.37	.38	14,8
Rimes 3 min c	.54*	.55*	.70**	.44	14,8
Onsets 3 min c	.52	.15	.49	.73	11,7
K-4 science word matching 3 min c	.25	.18	.47	.54	14,8
K-4 science word identification 3 min c	.60*	.70**	.88**	.42	14,8
Simple maze 3 min c	.50	.44	.71**	.83*	14,8
Mixed letter identification 3 min c	.35	.49	.69**	.25	15,8
Fry's 100 word identification 3 min c	.61*	.60*	.82**	.28	14,8

* = correlation significant at .05 level, ** = correlation significant at .01 level

PPVT SS = Peabody Picture Vocabulary Test standard score; W-J word id SS = Woodcock-Johnson word identification subtest standard score; Checklist II = Early Literacy Knowledge and Reading Readiness Checklist, Version II, MTAS C = Minnesota Test of Academic Skills correct with minimal assist.; c = correct; results are adjusted for prompting and guessing with a 3 consecutive error rule

Figure 1. Functional pictures and signs

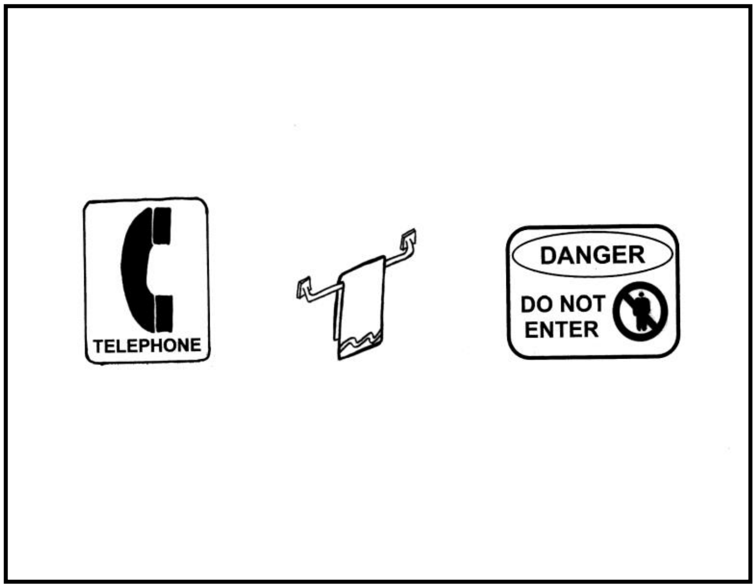


Figure 2. K-4 science content word identification

flexible

positive

wind

Figure 3. Rimes

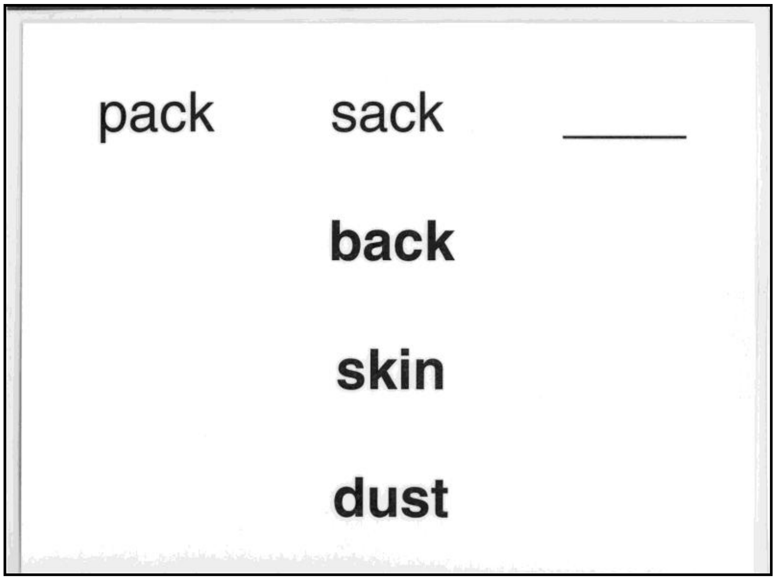
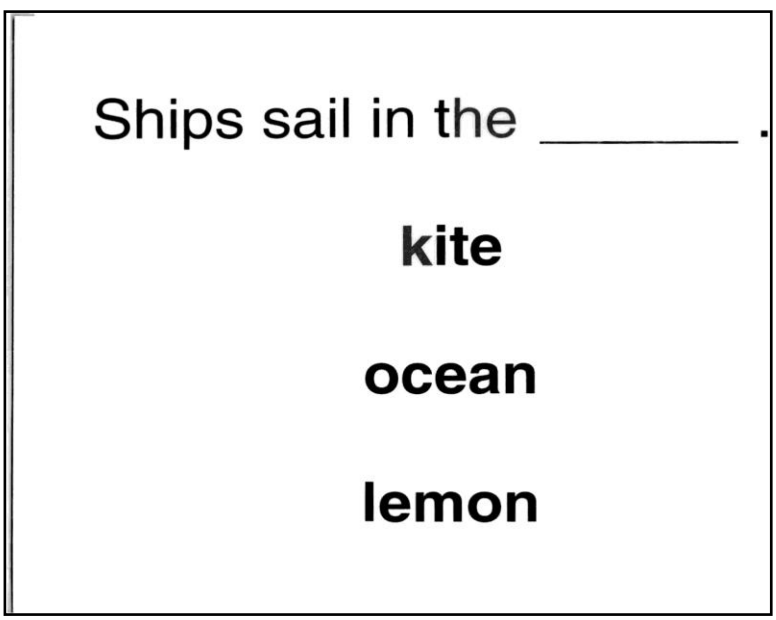


Figure 4. Simple maze



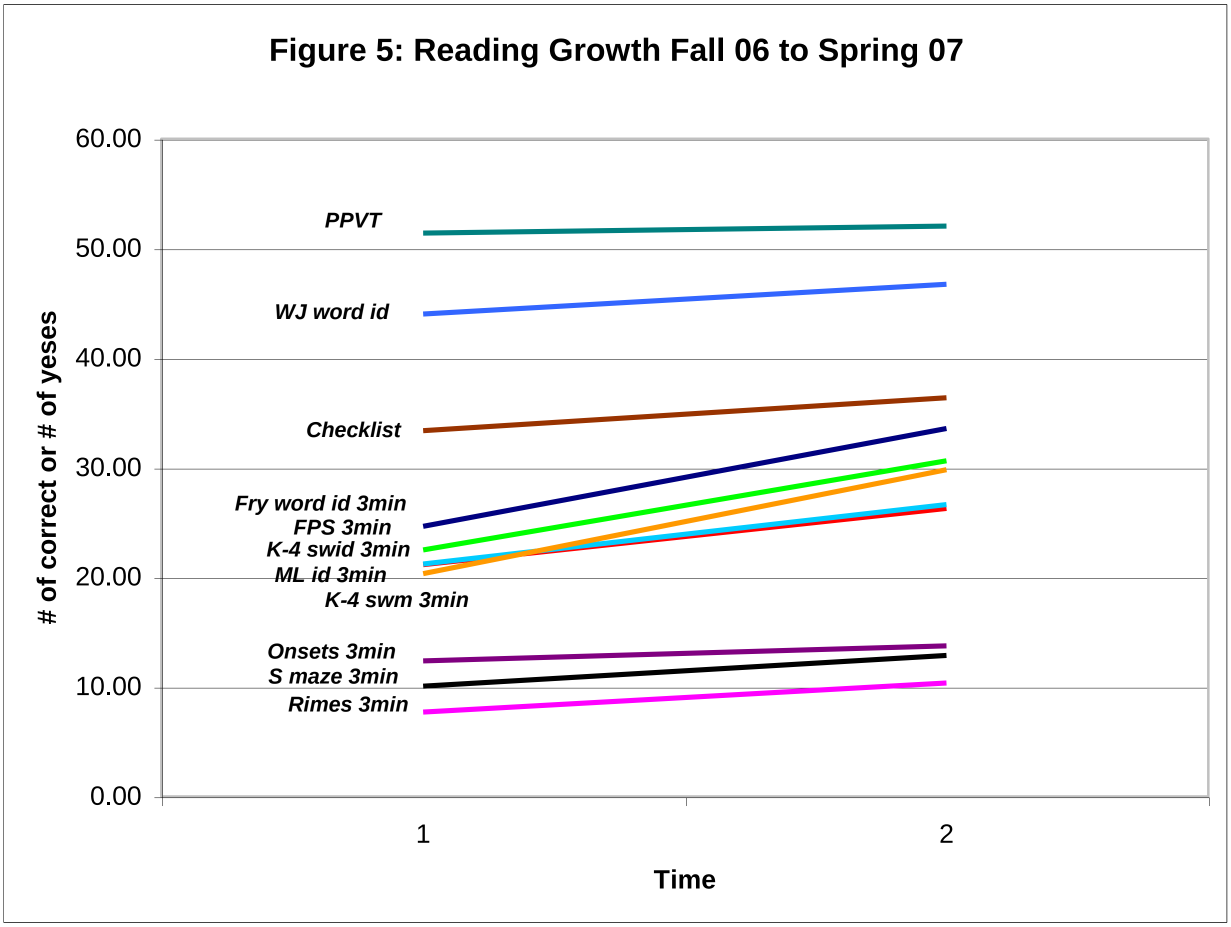


Figure 5: Reading Growth Fall 06 to Spring 07
60.00
50.00
40.00
30.00
20.00
10.00
0.00
# of correct or # of yeses
PPVT
WJ word id
Checklist
Fry word id 3min
FPS 3min
K-4 swid 3min
ML id 3min
K-4 swm 3min
Onsets 3min
S maze 3min
Rimes 3min
1
2
Time