Research Institute on Progress Monitoring

TECHNICAL REPORT #37:

Technical Characteristics of General Outcome Measures (GOMs) in Mathematics for Students with Significant Cognitive Disabilities

Renáta Tichá and Teri Wallace

RIPM Years 4 and 5: 2007 – 2008

Date of Study: September 2007 – May 2008

January 2010

The College of Education & Human Development

University of Minnesota

IDEAs that Work

U.S. Office of Special Education Programs

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 purpose of this two-year study was to examine the reliability, validity, and sensitivity to growth of newly developed general outcome measures (GOMs) in mathematics for teachers to use with students with significant cognitive disabilities. General outcome measurement framework, existing research in early mathematics education using this framework as well as the knowledge of educational needs of students with significant disabilities served as a basis for this study. The participants were 26 students with significant cognitive disabilities ranging from 1^{st} to 10^{th} grade. Technical characteristics of four new GOMs were examined, Number Identification (NI), Number Order (NO), Quantity Discrimination (QD), and Number Facts (NF). In year one, NI, NO, QD were administered for 5 minutes each. Based on the results from the first year, NI, NO, and NF were administered in the second year for 3 minutes each. Records were also kept for 1 minute timings. Results revealed that the new GOMs can be used reliably and NI, NO, and NF have promising concurrent (.51 to .79, $p < .01$) and predictive (.41, $p < .05$, to .73, $p < .01$) validity with the Early Math Diagnostic Assessment (EMDA) and the RIPM Early Numeracy Knowledge and Math Readiness Checklist (Math Checklist). The results of longitudinal data analyses demonstrated sensitivity to growth of NI and the Math Checklist that was significantly related to student initial performance on the EMDA. The results of this study indicate the potential of general outcome measures for students with significant cognitive disabilities in math. Further research is needed.

Technical Characteristics of General Outcome Measures (GOMs) in Mathematics for Students with Significant Cognitive Disabilities

Introduction

Educational accountability highlighted by the No Child Left Behind act (NCLB) of 2001 has increased the need for states, districts, schools, and teachers to be aware of how their students are progressing in academic areas. The results of state assessments administered at the end of each school year serve as an indicator of annual yearly progress (AYP) for each school under NCLB. Thus, performance on these state assessments is highly scrutinized not only at national, state and district levels, but also by parents and communities at large. Although state assessments have very high stakes, there is very little that the tests themselves can provide to inform teachers throughout the school year about how their students are progressing. State tests represent a summative approach to assessment, where academic performance is evaluated as a whole at one point in time. However, having knowledge about students' progress (i.e., using the results of a formative assessment) would help teachers focus their instruction accordingly.

One effective system of formative assessment, often referred to as progress monitoring, is curriculum-based measurement (CBM). Curriculum-based measurement was developed by Stanley Deno and his associates at the Institute for Research on Learning Disabilities (IRLD) in the late 1970s and early 1980s at the University of Minnesota (Deno, 1985; Marston, 1989). Over the last 30 years, extensive research has demonstrated that CBM is an assessment with sound technical characteristics (see reviews by Marston, 1989; and Wayman, Wallace, Wiley, Ticha, and Espin, 2007;

Foegen, Jiban, and Deno, 2007; McMaster and Espin, 2007). Curriculum-based measurement was developed in the context of special education for teachers to collect, score, graph and visually examine student data to modify instruction in order to improve student achievement, initially in elementary reading. Based on the research behind the elementary reading CBM measures and their subsequent use in the classroom, CBM measures in other areas have been developed (e.g., early literacy, writing, mathematics).

Progress Monitoring Measures in Mathematics

Progress monitoring in mathematics has been characterized in terms of two approaches to developing stimulus materials: the curriculum-based approach and the general outcomes approach (Foegen, Jiban, and Deno, 2007). Foegen et al. (2007) conducted an extensive review of literature on progress monitoring in mathematics for students from preschool to secondary school. Current research in progress monitoring measures in mathematics has focused primarily on typically developing students and students with mild disabilities. Foegen et al. (2007) reviewed progress monitoring math measures from pre-K to secondary level. Only the math measures from pre-K to 1st grade will be highlighted here because developmentally they are most relevant to students with significant cognitive disabilities (Griffin, 2003).

In early mathematics, the research is limited and has only focused on typically developing students. Unlike in later grades, early math measures fall under the category of general outcome measures, assessing early numeracy. The researched early numeracy measures include Quantity Discrimination, Number Identification, and Identifying the Missing Number in a counting sequence for students in pre-K, K and 1st grade (Chard, Clarke, Baker, Otterstedt, Braun, and Katz, 2005; Clarke and Shinn, 2004; in Foegen et al., 2007), and Circling Numbers, Drawing Numbers and Drawing Circles for students in pre-K and K (VanDerHeyden, Broussard, Fabre, Stanley, LeGendre, and Creppell, 2004; VanDerHeyden, Witt, Naquin, and Noell, 2001; in Foegen et al., 2007).

Both groups of researchers examined two additional early numeracy measures: Number Naming/Identification and Counting Tasks. Except for one measure called Choose Shape that showed a lower reliability coefficient (.40), reliability of the early numeracy measures ranged from .70 (“draw circles”) to .99 (“oral counting”). Foegen et al. (2007) found a greater spread in criterion validity coefficients among the early numeracy measures in the four studies examined. As in the case of reliability, the Choose Shape measure produced the lowest criterion validity coefficient (.06). In general, Foegen et al. (2007) reported that the early numeracy measures examined by VanDerHeyden and colleagues produced lower criterion validity coefficients than (from .06 for Choose Shape to .61 for Circle Number) than the measures examined by Clarke and colleagues (from .49 for Oral Counting to .80 for Quantity Discrimination). Only Clarke and colleagues investigated sensitivity of the early numeracy measures to growth (Foegen et al., 2007). The Number Identification measure in the study by Chard et al. (2005) detected the greatest improvement over 32 weeks in both K and 1st grade. In contrast, in the study by Clarke and Shinn (2004), Oral Counting demonstrated the greatest growth in 26 weeks, followed by Number Identification for 1st-grade students.

In their review, Foegen et al. (2007) report on six studies that included 1st-grade students along with older students in their sample when examining various computation measures. Three studies report results on 1st-grade students specifically. VanDerHeyden, Witt, and Naquin (2003) report test-retest reliability of .95 for their addition measure administered to 1st-graders. Fuchs, Fuchs, Hamlett, Waltz, and German (1993) and Shapiro, Edwards, and Zigmond (2005) report slope values for 1st-graders using the most established computation measures, the MBSP Computation, developed by Fuchs, Hamlett, and Fuchs (1998) as .53 and .32 respectively.

Since the review by Foegen et al. (2007), new studies have been published investigating early numeracy measures for K and 1st-grade students (Clarke, Baker, Smolkowski, and Chard, 2008; Martinez, Missall, Graney, Aricak, and Clarke, 2008; Methe, Hintze, and Floyd, 2008; and Lembke, Foegen, Whittaker, and Hampton, 2008). Clarke et al. (2008) investigated whether slope adds to predictive accuracy beyond information gained from a static performance score at the beginning of a school year for Oral Counting, Number Identification, Quantity Discrimination, and Missing Number for 254 K students. Predictive validity between the early numeracy measures and the Stanford Early School Achievement Test (SESAT) ranged from .55 for Oral Counting to .60 for Quantity Discrimination. Only growth on Quantity Discrimination explained additional variance on the SESAT.

A study by Martinez et al. (2008) focused on examining the technical adequacy of Oral Counting, Number Identification, Quantity Discrimination, and Missing Number with 59 K students. Delayed alternate-form reliability ranged between .77 for Quantity Discrimination to .91 for Number Identification. Test-retest reliability ranged from .80 for Quantity Discrimination to .92 for Number Identification. The Quantity Discrimination measure demonstrated the best concurrent validity with the SAT-10 administered to the K students in the spring (.64). All CBM measures administered in the fall were significantly related to students' performance on the SAT-10 in the spring (.46

for Quantity Discrimination, .45 for Oral Counting, .36 for Missing Number, and .31 for Number Identification). In addition, Martinez et al. (2008) investigated growth from fall to spring on all the CBM measures except Oral Counting. All three measures detected significant growth over the period of 28 weeks with an average weekly growth of .46 correct responses for Number Identification, .32 for Quantity Discrimination, and .24 for Missing Number.

Methe et al. (2008) investigated four early CBM math measures they referred to as Early Numeracy Skill Indicators for use with 64 K students: Counting-on Fluency (COF), Ordinal Position Fluency (OPF), Number Recognition Fluency (NRF), and Match Quantity Fluency (MQF). Test-retest reliability ranged between .74 for MQF and .98 for NRF. Validity of the newly created measures was established using the Test of Early Mathematics Achievement (TEMA-3) and teacher ratings. Concurrent validity in the fall ranged between .50 for COF and .72 for NRF with TEMA-3 and between .68 for COF and .89 for NRF with teacher ratings. In the spring, MQF demonstrated the lowest validity with both TEMA-3 and teacher ratings (.20 and .66 respectively), while NRF the highest (.64 and .89 respectively). Fall to spring predictive validity with TEMA-3 ranged between .41 for MQF and .70 for NFR and between .57 for COF and .87 for NRF with teacher ratings.

Lembke et al. (2008) examined the sensitivity to growth of Quantity Discrimination, Missing Number, and Number Identification measures for 77 K and 30 1st-grade students across 28 weeks on a monthly basis. Lembke et al. (2008) found that K as well as 1st-grade students demonstrated significant linear growth on the Number Identification measure. The estimated weekly slope for K students was .34 and .24 for

1st-grade students. The growth on Quantity Discrimination and Missing Number was curvilinear.

It is clear from the studies reviewed that research is lacking on progress monitoring measures for students with significant cognitive disabilities as is indicated by the fact that none of the studies examining early math measures in the review by Foegen et al. (2007) included students receiving special education services. The research conducted on students in pre-K, K and 1st grade can serve as guidance for the development of general outcome measures for students with significant cognitive disabilities.

Mathematics Instruction for Students with Significant Cognitive Disabilities

For the purposes of this review, students with significant cognitive disabilities are defined as those who take the alternate assessment based on alternate achievement standards. In the language of the Individuals with Disabilities Education Act 2004 (IDEA), the focus of the review is on students with mild and moderate mental retardation (AAMR, 2002).

The review of research on progress monitoring measures in pre-K, K and 1st grade revealed that students with significant cognitive disabilities have not been included in those studies. The results of an assessment in a particular area should reflect what a student has learned. Examination of technical characteristics of an assessment or a measure gives the assessor confidence that the results of the assessment reflect materials learned. It is typical in a general education classroom to use a curriculum with a sequence of skills and content to be taught in an academic area.

In mathematics, the National Council of Teachers of Mathematics (NCTM) in 2000 outlined the “Principles and Standards for School Mathematics” to be followed when teaching. The standards fall into two general categories, content and process. The content standards include: numbers and operations, measurement, data analysis and probability, geometry, and algebra. The process standards include: problem solving, reasoning and proof, connections, communication, and representation. Syllabi and curricula in general education classrooms across grade levels reflect these outlined standards. Even though inclusion of students with disabilities in general education has never been more encouraged than under the No Child Left Behind act (NCLB), reviews of research have shown that instruction for students with significant cognitive disabilities in mathematics has not followed the general education standards as outlined by NCTM (Browder and Spooner, 2006; Browder, Spooner, Ahlgrim-Delzell, Harris, and Wakeman, 2008).

Browder et al. (2008) conducted an extensive review and meta-analysis of 68 experimental studies (54 were single subject and 14 were group designs) examining mathematics instruction for 493 students with significant cognitive disabilities. Fifty six percent of the instruction took place in special education classrooms, 26 percent in the community, 35 percent in general education classrooms, 13 percent at home, 4 percent in employment settings, and another 4 percent in residential facilities (16 studies took place in multiple settings). Browder et al. (2008) found that the majority of the studies addressed the two content standards targeted for younger students, numbers and operations (37 studies) or measurement (36 studies). Only two studies addressed each algebra, geometry, and data analysis. Out of the numbers and operations standard studies, 12 examined calculation, nine matching numbers, nine counting, and seven additional types of instruction in mathematics. Money skills instruction was the object of study of 33 out of the 36 measurement studies, while time instruction only of three measurement studies. Browder et al. (2008) concluded that students with significant cognitive disabilities are capable of learning skills under the standards numbers and operations as well as measurement. Browder et al. (2008) also noted that there is not a sufficient number of studies that fall under algebra, geometry and data analysis to make a conclusion about whether students with significant cognitive disabilities can successfully learn skills under those three mathematics standards.

The results of the literature review and meta-analysis by Browder et al. (2008) is likely to be a reflection of instructional practices used when teaching students with significant cognitive disabilities. Targeting at least some aspects of all mathematics standards as outlined by NCTM and prioritizing the skills for each student within those standards should be a goal of educators of students with significant cognitive disabilities. At the same time, more research is warranted to empirically support the benefits of this approach (Browder and Spooner, 2006; Browder et al., 2008). Moreover, Browder and Spooner (2006) described in detail along with examples how this approach can and should be balanced with instruction aligned with IEP goals in mathematics for each individual student.

Mathematics Assessments for Students with Significant Cognitive Disabilities

The content students with significant disabilities are taught in the classroom or other settings is tied more or less directly to the subsequent assessment of that content. Different types of assessment have stronger or weaker links with the content taught.

Commercially developed achievement tests typically have a weaker link with the instructional content than assessments developed by teachers themselves (e.g., mastery monitoring of a skill taught or portfolio assessment). Commercially developed achievement tests tend to assess a general achievement level of students in a content area. Moreover, in math there is a lack of commercially developed achievement tests designed specifically for students with significant cognitive disabilities. In a review of 27 norm-referenced aptitude and achievement tests, Fuchs, Fuchs, Benowitz, and Barringer (1987) found that most of the tests reviewed have not provided information on including students with disabilities in the normative sample. Consequently, Fuchs et al. (1987) concluded that the norms did not reflect performance of students with disabilities. Even though the study by Fuchs et al. (1987) dates back 20 years, the trend has not changed markedly up to date.

In contrast to commercially developed achievement tests in mathematics, teacher-developed assessments tend to have a closer connection to the material taught in the classroom. Among the most widely used teacher-developed assessments to evaluate learning of students with significant cognitive disabilities are mastery monitoring and portfolios. Mastery monitoring refers to assessing mastery of a particular skill (e.g., money skills). A portfolio is an assembly of a student's work in an academic area over time. Mastery monitoring and portfolio assessments have a greater instructional value for teachers and students than commercially developed tests because they track student progress of a skill or performance in a certain area. However, teacher-developed assessments have a different disadvantage, undetermined technical characteristics. In addition, because mastery monitoring is limited only to mastering a particular skill or a limited content, it does not help the teacher see student progress more broadly across a whole curriculum or a general outcome.

Since NCLB came into effect, students with significant cognitive disabilities participate in alternate assessments in mathematics in grades 3–8 and at least once in grades 10–12 (Briggs, 2005). Types of alternate assessments differ by state. Most typically, alternate assessments have been reported as being in the form of portfolios, performance assessment, or rating scales. Alternate assessments need to be aligned with the state's content standards (Elliott and Roach, 2007). Similar to commercially developed tests, alternate assessments have limited instructional value, primarily because of their infrequent administration (i.e., once a year), or because alternate assessments have not been developed with the same psychometric rigor as commercially developed tests, which limits the reliability and validity of their results (Perner, 2007). Alternate assessments have been evolving following NCLB regulations and since 2001 (Elliott and Roach, 2007).

In Minnesota, the alternate assessment in mathematics is based on alternate achievement standards that are directly aligned with the Minnesota content standards (Minnesota Department of Education, 2008). The assessment is administered by special education teachers at the end of the school year. The Minnesota alternate assessment is a multiple-choice test designed to sample student knowledge without having to cover every standard. The assessment is scored with a 4-point rubric. In 2008, Minnesota special education teachers administered a second version of the alternate assessment with the intent to provide more reliable and valid results than the first administration.

In order to overcome the shortcomings of commercially developed, teacher developed and state developed assessments in mathematics for students with significant cognitive disabilities, namely the appropriateness of use, technical rigor and instructional utility, another line of assessment that has traditionally been used with students with mild disabilities has begun to be explored.

Progress Monitoring in Mathematics for Students with Significant Cognitive Disabilities

An approach to assessment developed by researchers in special education, curriculum-based measurement (CBM; Deno, 1985) or general outcome measurement (GOM) in early childhood (McConnell, McEvoy, Carta, Greenwood, Kaminski, Good, and Shinn, 1998), is designed to sample from a curriculum or across general skills across time to monitor student progress (see Introduction). Although CBM was developed in the context of special education, it was developed for use with students with mild and learning disabilities. From the review of progress monitoring research in mathematics by Foegen et al. (2007), it is clear that progress monitoring measures for students with significant cognitive disabilities are lacking.

Browder, Wallace, Snell, and Kleinert (2005) in a white paper for the National Center on Student Progress Monitoring have outlined three unique challenges for progress monitoring of students with significant cognitive disabilities. First, students with significant cognitive disabilities may have unique ways to respond to content and assessment materials due to their disability. Second, the traditional focus of instruction for students with significant cognitive disabilities has been on functional skills (e.g., money skills), which does not lend itself to progress monitoring across a curriculum (CBM) or a general outcome (GOM). Third, there are no guidelines on what progress should be expected from students with significant cognitive disabilities within a general curriculum in each content area.

The purpose of this study was to begin developing progress monitoring measures in mathematics for students with significant cognitive disabilities. In order to overcome the challenges outlined by Browder et al. (2005), our goal was to create measures that did not require a verbal response, represented general mathematical skills, and were based on the basic principles of CBM (i.e., to provide teachers with a reliable, valid, simple, efficient, easily understood, and inexpensive alternative to commercial standardized tests and informal observations for monitoring student progress; Deno, 1985).

Method

There were three main research questions posed by this study:
Are the math GOMs (1) reliable, (2) valid and (3) sensitive to growth over time when used with students with significant cognitive disabilities?

Participants

The participants in this two-year study were 26 students with significant cognitive disabilities from an urban school district in Minnesota. Nineteen students (73%) were male and seven (27%) were female. Students from grade one through grade ten were represented (two 1st-grade, two 2nd-grade, four 3rd-grade, six 4th-grade, four 5th-grade, one 6th-grade, four 7th-grade, one 9th-grade, and two 10th-grade students). There were 11 (42.3 %) African American, five (19.2%) Hispanic, nine (34.6%) White, and one (3.8%) Native American students. Twenty-one of the 26 students (80.8%) received free or reduced lunch. Four students (15.4%) were English Language Learners (ELL). Based on information in their IEPs, the primary disability of the students was as follows: DCD (developmental cognitive disability) was a primary label of 16 students (61.5%), two students were labeled SMI (severe multiple impairment, 7.7%), one OHI (other health impairment, 3.8%), one TBI (traumatic brain injury, 3.8%), and six students had a label specific to the district, SNAP (student needing alternative program, 23.1%). In the case of the students whose primary disability label was not DCD, their secondary or tertiary label suggested this impairment.

The demographic characteristics 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). The comparison between the sample in this study and the school district demographic composition demonstrates a good representation in the category of ELL. Male students were over-represented in our sample, as were students receiving free or reduced lunch. In the category of ethnicity, African-American and Native American students were under-represented in our sample, while Hispanic and White students were over-represented.

Measures

Four general outcome measures (GOM) in mathematics were examined in this study: Number Identification (NI), Number Order (NO), Quantity Discrimination (QD), and Number Facts (NF). Each GOM consisted of a set of 60 laminated 8.5x11 inch cards. Three out of the 60 cards were practice cards and 57 cards were test cards. None of the math GOMs required a verbal response. Number Identification (NI), NO, and QD were administered in year 1 for the duration of 5 minutes. Based on the results in year 1, in year 2 NI, NO and NF were administered for 3 minutes only.

Each Number Identification (NI) card contained three different randomly selected numbers in the range from 0 to 100. Students were asked to point to the number verbalized by the administrator. Each Number Order (NO) card contained two boxes with unequal numbers ranging from 0 to 100. Students were asked by the administrator to point to the box with a greater number. Each Quantity Discrimination (QD) card contained two boxes with an unequal number of dots ranging from 0 to 20. Students were asked to point to the box representing a greater quantity. Each Number Facts (NF) card contained an addition or a subtraction problem at the top of the card using numbers 0 to 10. Below the problem were three different number choices ranging from 0 to 10 as answers. Students were asked to point to one of the numbers that represented the correct answer to the problem. Each test card was presented to students for 5 seconds or less, depending on the speed of their response. See Figures 1-4 for examples of the math GOMs.

In addition, three criterion measures were used: the Early Math Diagnostic Assessment (EMDA; The Psychological Corporation, 2002), the RIPM Early Numeracy Knowledge and Math Readiness Checklist – Version II (Math Checklist; Research Institute on Progress Monitoring, 2007), and the math portion of the Minnesota Test of Academic Skills (MTAS; Minnesota Department of Education, 2008). The EMDA is a norm-referenced standardized assessment of math skills designed for use with students in K – grade 3. The randomly stratified norming sample included 1,374 students in PreK through grade 3 and was collected in the 1999-2000 and 2000-2001 school years. Students receiving special education services were included in the sample. Six to 10% of the sample were students with a learning disability, speech/language impairment, emotional disturbance, mild mental impairment, attention deficit disorder, or a physical impairment. It takes 15 – 20 minutes to administer and is administered individually. The EMDA is aligned with the Principles and Standards for School Mathematics (NCTM, 2000). The assessment has two subtests: Math Reasoning and Numerical Operations. The Math Reasoning subtest requires pointing and verbal responses, while the Numerical Operations subtest requires fine motor responses, e.g. circling or writing numbers. The skills assessed with EMDA range from early numerical concepts to quantitative and qualitative applications (The Psychological Corporation, 2002). Test-retest reliability reported was .92 for Numerical Operations and .96 for Math Reasoning. Criterion validity coefficients for Math Reasoning ranged from .67 with the corresponding subtest of the Wide Range Achievement Test – III to .82 with the corresponding subtest of the Wechsler Individual Achievement Test. Validity coefficients for Numerical Operations ranged from .75 with the corresponding subtest of the Wide Differential Ability Scales to .78 with the corresponding subtest of the Wechsler Individual Achievement Test.

Due to the lack of normed and standardized assessments in math appropriate for students with significant disabilities, the Math Checklist was developed by the researchers. It was believed that special education teachers of these students would have an especially acute knowledge of their performance based on the smaller number of students special education teachers work with on regular basis in comparison to general education teachers and a detailed monitoring of student performance based on the their IEPs. The Math Checklist targeted teacher judgment of their students' math skills in five areas: I. Number Sense, II. Computation, III. Space, Shape, and Measurement, IV. Functional Math, and V. Calculator Skills. The school district DCD Scope and Sequence document in Math, the district Early Childhood Special Education Checklist, and conversations with special education teachers served as bases for developing the Math Checklist. Teachers recorded a positive response if they judged a student as proficient in a particular math subskill (e.g., “Can multiply single digit numbers”), and a negative response if a student had not mastered the particular subskill yet. The number of yes and no responses was recorded for each subscale as well as total scores. The predictive validity results of this study indicated an $r = .81$ ($p < .01$) relationship between the Math Checklist and the EMDA.

The MTAS is an alternate assessment for students with the most significant cognitive disabilities in Minnesota. It is a statewide assessment based on alternate achievement standards. The test is individually administered by a special educator and is untimed. The MTAS in math is administered in the spring in grades 3-8 and 11. The purpose of the MTAS is to sample student knowledge without having to assess students on every standard or benchmark. The MTAS in math consists of 15 tasks under four strands: 1. Number Sense, 2. Patterns, Functions and Algebra, 3. Data Analysis, Statistics and Probability, and 4. Spatial Sense, Geometry and Measurement. The test is scored using a scoring rubric on a scale from 0-3 that reflects the independence and correctness of the student’s response (Minnesota Test of Academic Skills, 2008). Because the MTAS is a newly developed assessment, state-wide reliability or validity data have not yet been reported.

Procedures

General outcome measures (GOMs) development. To be able to fulfill the purpose set for this study, we combined the knowledge generated by research in early numeracy progress monitoring with the knowledge generated by research on instruction and assessment of students with significant cognitive disabilities in general and mathematics in particular. By combining these two sets of research-generated knowledge, it was important to assure both that the proposed assessments tasks be meaningful to students with significant cognitive disabilities (Browder and Spooner, 2006) and that the tasks be aligned with content standards in mathematics (Perner, 2007).

Because the development of general outcome measures (GOM) in mathematics for students with significant cognitive disabilities is in its initial stage, we began with the first mathematics content standard, “numbers and operations”. Important components of this standard are number identification, one-to-one correspondence, understanding place value, and number order, and other aspects of number sense (e.g., addition and subtraction; Browder and Spooner, 2006). This approach was followed in creating the early numeracy measures. Number Identification (NI), Quantity Discrimination (QD), and Number Order (NO) measures were adapted from the study by Clarke and Shinn (2004) where they were initially used with typically developing students in 1st grade. The NI measure used by Clarke and Shinn (2004) was modified from a measure requiring a verbal response to a measure requiring a pointing response only, by adjusting its duration from 1 to 5 minutes in year 1 and to 3 minutes in year 2 of the study, and by limiting presentation to one problem per page. The QD measures in Clarke and Shinn (2004) was modified into the NO measure in this study using the same three modifications as in NI.

The QD measure in this study was created from a Quantity Array measure in Lembke and Foegen (2007) and the QD measure in Clarke and Shinn (2004) using numbers rather than dots. In the study by Lembke and Foegen (2007), students were to identify the number of dots presented in boxes in 1 minute. Although in their study, Lembke and Foegen did not find Quantity Array to have strong technical adequacy characteristics for students in K and 1st grade, our intent was to create a pre-symbolic measure for quantity discrimination effective for students with significant cognitive disabilities. As in the NI and NO measures created for this study, the QD measure required pointing responses only, was timed for 5 minutes in year 1 and 3 minutes in year 2 of the study, and presented students only with one problem at a time.

The NF measure was created in year 2 of the study based on the weak technical adequacy results of the QD measure in year 1 and the need for a measure requiring more complex math operations for more cognitively advanced students. The NF measure was adapted from 1st-grade probes of a well established computation measure, the MBSP Computation, combining addition and subtraction operations using numbers 0-100 (Fuchs, Hamlet, and Fuchs, 1998). The adaptations in creating the NF measure followed the three adaptations applied when creating the NI, QD, and NO measures to better serve students with significant cognitive disabilities.

Measure administration and data collection. The data for this study were collected across two years in winter 07, fall 07, and spring 08. Each GOM was administered individually by two primary data collectors. Secondary data collectors were present for approximately 80% of GOMs administration and for 40% of EMDA administration. The primary data collector administered the model card, two practice cards, and test cards one by one to each student according to standardized directions, while recording student responses on a scoring sheet. The secondary data collector's role was to shadow the first data collector in recording student responses, time intervals, and other observations for accuracy, and to generally assist the primary data collector. Both sets of data collectors used a small portable tape recorder with an ear piece to be able to accurately record student responses at 1, 3, or 5 minute intervals as well as a timer to monitor student response time to each card administered for 5 seconds. Data collectors were graduate students in education or educational psychology. All data collectors participated in a training session prior to data collection that addressed not only the administration procedures of the measures but also working in the schools with students with significant cognitive disabilities.

Two out of the four math GOMs (NI and NO) were administered across all three data collections in winter 07, fall 07, and spring 08. The Quantity Discrimination (QD) measure was only administered in winter 07, while the Number Facts (NF) measure was administered in fall 07 and spring 08. The same form of each GOM was administered across time. The GOMs were administered to students in a counterbalanced order by randomly assigning students to four different orders of the measures (math GOMs were administered along with reading GOMs). The EMDA, in the same form, was administered to students twice in the course of the study, in winter 07 and winter 08. The Math Checklist was completed by teachers three times in the study, in winter 07, fall 07 and spring 08. Two forms of the Minnesota alternate assessment (MTAS) not compatible with each other were administered to students with significant cognitive disabilities district-wide in spring 07 and spring 08.

Additional data on the students in the study were collected from the district database in the form of demographic information, such as grade, SES, primary disability, and ELL status, as well as IEP goals and objectives in math. The Minnesota Test of Academic Skills (MTAS) scores were also obtained from the district.

Prompting system. A four-level prompting system was applied when assessing students with GOMs in this study in order to ensure that all students were able to respond to the items on the measures. Level 0 represented non-prompted responses, Level 1 prompts were verbal prompts repeating directions 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. In addition, in winter of year 1, behavioral directives were recorded along with levels of prompts in case a student needed to be redirected to task.

Scoring. Student responses on the GOMs were recorded on a scoring sheet common across all GOMs. On the front page, the data collectors recorded “0” for an incorrect or “1” for a correct response on two practice items along with a level of prompt the student needed to give a correct response (0, 1, 2, or 3). On the next 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. Only 2% of student responses were aided with a prompt during GOM testing. In the analyses, those responses that required a prompt were treated as incorrect. The prompting system provided an opportunity for all students in the study to respond to the practice and test items and gave students practice to be able to respond to the items independently in the future.

Scores on all GOMs were corrected for guessing using a 3-consecutive-error rule. According to this rule, the only scores counted as correct were those 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 and Lehman, 1991) conducted by Wallace, Ticha, and Gustafson (2009).

The Math Checklist was scored by counting the number of "yes" and "no" responses for each of the subscales and in total. The total number of positive responses was used for analyses. The EMDA was scored according to the standardized published directions. EMDA raw scores were used for analyses. Reported state math test (MTAS) scale scores were used in the study.

Analyses

First, data were analyzed using descriptive statistics in the form of means, standard deviations, and range of scores for the GOMs, Math Checklist, EMDA, and the math portion of MTAS. Even though frequencies of prompting levels and behavior directives implemented are reported, inferential analyses were computed on data corrected for both of those procedures. Inferential statistics in the form of Pearson correlations (Howell, 2002) were used to calculate test-retest reliability for all the measures, except the MTAS administered by the district. Pearson correlations were also computed between the math GOMs and all criterion measures in order to establish criterion validity for the newly developed GOMs. Hierarchical linear modeling (HLM; Raudenbush and Bryk, 2002) for repeated measures was used to examine growth over time, the relationship between growth and initial status, and whether growth was predicated by the EMDA total raw score (EMDA). The GOMs with three data points were examined separately with HLM: Number Identification 1 minute (NI1), Number Identification 3 minutes (NI3), Number Order 1 minute (NO1), Number Order 3 minutes (NO3), and the Math Checklist.

Results

Descriptives

Descriptive statistics of cross-sectional GOM data in year 1 and 2 revealed several important findings (Tables 1 and 2 in the Appendix). First, the initial mean scores and standard deviations of each math GOM at 1 minute indicated that the GOMs were at different difficulty levels. The following order of GOMs demonstrates their range of difficulty, starting from the most difficult to the easiest as indicated by the average score and the spread of cards identified correctly at 1 min respectively: Number Facts (NF) administered only in year 2, Number Order (NO) administered initially in year 1 , Number Identification (NI) administered initially in year 1 and Quantity Discrimination (QD) administered in year 1 only . The same pattern of ordering measures by difficulty held at the 3-min time frame. In addition, the spread of scores for the two hardest GOMs (NF and NO) was the largest in relation to the GOM means.

Second, the mean scores for each time frame in year 1 (1, 3, and 5 minutes) revealed greater average growth between 1 and 3 minutes than between 3 and 5 minutes for all GOMs (Table 1). This indicates that a 3-minute time limit for math GOMs for students with significant cognitive disabilities may be preferred. It needs to be noted, however, that one student at 3 minutes and four students at 5 minutes reached the ceiling on NI, while one student at 3 minutes and three students at 5 minutes reached the ceiling on QD.

Third, having examined the skew of all GOMs at all time frames, all distributions can be considered symmetrical ($\gamma_1 < + \text{ or } - 1$), except NO at 3 and 5 minutes in year 1and NF at 3 minutes in the spring of year 2 all with positively skewed distributions. All criterion measures were symmetrically distributed. Values describing kurtosis of measure distributions indicated normality for all measures.

Fourth, all measures, except for QD with data from only one data collection in year 1, demonstrated increase in mean scores across time (Figure 5). Student mean performance increased incrementally on measures with three data points in winter 07, fall 07, and spring 08. Number Identification (NI) mean correct scores at 1 and 3 minutes increased by .44 and 2.84 respectively. Number Order (NO) mean correct scores at 1 and 3 minutes increased by 2.06 and 4.67 respectively. The mean scores for Math Checklist increased by 8.04 items answered positively by the teachers. Two measures, NF and the EMDA were administered only twice in the course of the study. The mean correct scores of NF administered in the fall and spring of year 2 increased by .44 in 1 minute and by .29 in 3 minutes. The EMDA total correct score increased from by 1.84 from winter 07 to winter 08 It should be noted that for the GOMs the increase in mean correct scores was greater at a 3- rather than 1-minute time frame, except for the most difficult NF.

Reliability

Two types of reliability were calculated to evaluate the technical adequacy of GOMs: inter-observer and inter-scorer. Reliability was calculated by dividing the larger score by the smaller. The scores calculated from the scoring sheets of the data collector and observer on the same student were compared for inter-observer reliability. Inter-observer reliability was calculated, on average, on 53% of GOMs and 36% EMDA assessments across years 1 and 2. In year 1, there was a 100% inter-observer agreement on all the measures. In year 2, there was a 100% agreement on all the measures in the spring and all the measures but NF (99%) in the fall. The scores calculated from the same scoring sheet by two different scorers were compared for inter-scorer reliability. Inter-scorer reliability was calculated, on average, on 24% of GOMs, 23% of Math Checklists, and 28% of EMDA assessments across the two years. In year 1, the average inter-scorer reliability was 98%, ranging from 96% for QD to 100% for NI. The inter-scorer reliability for the Math Checklist in year 1 was 99% and 100% for the EMDA. In year 2, the average inter-scorer reliability for the GOMs was 95% (range between 90% for NI and 100% for NF) in the fall and 99% (range between 98% for NF and 100% for NI and NO) in the spring. The inter-scorer reliability for the Math Checklist in year 2 was 99% in the fall and 100% in the spring. Inter-scorer reliability for the EMDA was again100%.

Validity

Two types of criterion validity were calculated for the purposes of this study: concurrent in year 1 and 2 and predictive in year 2. Table 3 displays the concurrent relationship between the GOMs at 1 and 3 minutes and two criterion measures, the EMDA and Math Checklist administered in years 1 and 2. The results indicate that NO at 1 and 3 minutes related to the criterion measures the strongest most consistently (correlations range between .60 and .71, $p < .01$), followed closely by NI (correlations range between .51 and .63, $p < .01$). Correlations with the EMDA at 1 and 3 minutes are similar, i.e. differ by .02 to .04 points across GOMs, while the differences for the Math Checklist the correlations between 1 and 3 minutes were larger (.05 to .10). Based on the lower correlations between QD and the criterion measures (range between .39 and .44, $p < .05$) in year 1, in year 2 this GOM was replaced with the more challenging NF. The results of concurrent validity in year 2 indicate that NF related to the Math Checklist the strongest (.78 and .79, $p < .01$). The correlations between NO and the Math Checklist increased in year 2 at 1 minute (.65, $p < .01$) and remained the same (.68, $p < .01$) at 3 minutes. In the case of NI, however, the relationship with the Math Checklist weakened by .11 at 1 minutes and .05 at 3 minutes but was still significant at $p < .01$. In year 2, the maximum difference between correlations at the two different time frames for GOMs was .03, and thus smaller than in year 1.

Results in Table 4 demonstrate the predictive relationship between GOMs administered in the fall of year 2 and the EMDA administered in winter and the Math Checklist in the spring of year 2. Correlations with NO follow a similar pattern to concurrent validity results indicating the strongest and most consistent relationship with criterion measures (range between .69 and .73, $p < .01$). The results for NI and NO are similar at 1 and 3 minutes. The correlation results for NF indicate that predictive validity differs from concurrent validity results in year 2. The correlation of NF at 1 minutes with the Math Checklist is lower. There is a wide difference between the correlations at 1 and 3 minutes with both the Math Checklist and the EMDA. The correlations at 1 minute are .15 and .17 respectively lower than the correlations at 3 minutes.

Validity results with the math MTAS scale scores are presented in Table 5. There is a clear distinction between the results of the three GOMs as related to the MTAS.

Based on their initial mean scores, the three GOMs have differentiated themselves by difficulty, with NI as the least difficult and NF the most difficult. The validity results with the MTAS appear to follow this descriptive differentiation. Except for the results from fall 07 at 1 minute, NF demonstrates the strongest relationship with the math MTAS (.51 - .54, $p < .05$). Number facts (NF) is followed by NO with non-significant correlations ranging from .36 to .46 and NI with non-significant correlations ranging from -.17 to .02. In addition, the relationship between the criterion EMDA and the MTAS was also non-significant at .30.

Growth

Hierarchical linear modeling (HLM) for repeated measures was used to examine growth over time, the relationship between growth and initial status, and whether growth was predicated by the EMDA total raw score (EMDA). The following variables were examined: Number Identification 1 minute (NI1), Number Identification 3 minutes (NI3), Number Order 1 minute (NO1), Number Order 3 minutes (NO3), and Math Checklist. For all analyses, time of year was coded as 0, 1, 2 and random intercepts and slopes were specified in the Level 1 model. The Level 2 model had a single static predictor, the EMDA.

Specifically, the Level 1 model was,

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij} \, ,$$

where Y_{ij} is the response score for the *i*th individual, $i = 1, \ldots, n$, at the *j*th time point, $j = 1, \ldots, n_I$, t_{ij} is the time score, β_{ki} is an individual-specific regression coefficient, and e_{ij} is a random error term. We assume the random error term is normally distributed with zero mean value and uncorrelated with any other terms.

The Level 2 model was,

$$\beta_{0i} = \beta_0 + \beta_2 EMDA_i + b_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_3 EMDA_i + b_{1i},$$

where β_k is a group-level regression coefficient and b_{ki} is a individual-specific random effect. We assume the random effects have a joint-normal distribution with zero means. In the Level 2 equation, β_2 indexes the strength of prediction of EMDA for the intercepts and β_3 indicates the strength of the prediction for the slope.

Parameters were estimated based on maximum likelihood methods. For each response variable the multi-parameter null hypothesis was first tested, H_0: $\beta_2 = \beta_3 = 0$, with a likelihood ratio chi-squared statistic to determine if EMDA had any effect. If the multi-parameter null hypothesis was rejected, then EMDA was left in the Level 2 model and *z*-tests for parameter estimates were computed. A median split of EMDA was used for graphing purposes but the full scale of the variable was used in the statistical analysis.

The estimated variance of the random slopes was very small in all cases. This resulted in a spurious positive perfect correlation between intercepts and slopes. The spurious correlation precluded accurate evaluation of the relationship between intercepts and slopes and here we focus on the nature of change over time and EMDA effects. Though the estimated variance of the random slopes was small, the term was left in the model because of observed variation in the graphs of the data presented below.

Number Identification 1 minute (NI1). Figure 6 shows the individual growth curves (thin lines) and mean growth curves (thick lines) over time of NI1 for the median split of EMDA (Low/High) and for the entire sample (all). Table 6 lists the results of the HLM analysis. The parameter estimates are in the upper portion and the multi-parameter test is in the lower portion. The table shows the multi-parameter test was statistically significant and the parameter estimates indicate there was a significant EMDA intercept effect. The positive sign of the β_2 parameter estimate indicates that higher EMDA scores were associated with higher initial starting values (i.e., higher intercepts). The positive sign of the β_1 parameter estimate indicate a slight increase in NI1 over time.

Number Identification 3 minutes (NI3). Figure 7 shows growth curves for NI3 and Table 7 lists the results of the HLM analysis. The table shows the multi-parameter test was statistically significant and the parameter estimates indicate there was a significant EMDA intercept effect. The positive sign of the β_2 parameter estimate indicates that higher EMDA scores were associated with higher initial starting values (i.e., higher intercepts). The positive sign of the β_1 parameter estimate indicate a slight increase in NI3 over time.

Number Order 1 minute (NO1). Figure 8 shows the individual growth curves for NO1 and Table 8 lists the results of the HLM analysis. The table shows the multi-parameter test was statistically significant and the parameter estimates indicate there was a significant EMDA intercept effect. The positive sign of the β_2 parameter estimate indicates that higher EMDA scores were associated with higher initial starting values (i.e., higher intercepts). However, unlike the previous results, the β_1 parameter estimate was relatively small and not statistically significant indicating a possibility of a zero population mean slope for the growth curves.

Number Order 3 minutes (NO3). Figure 9 shows the individual growth curves for NO3 and Table 9 lists the results of the HLM analysis. The table shows the multi-parameter test was statistically significant and the parameter estimates indicate there was a significant EMDA intercept effect. The positive sign of the β_2 parameter estimate indicates that higher EMDA scores were associated with higher initial starting values (i.e., higher intercepts). However, the β_1 parameter estimate was relatively small and not statistically significant indicating a possibility of a zero population mean slope for the growth curves.

Math Checklist (MC). Figure 10 shows growth curves for Math Checklist and Table 10 lists the results of the HLM analysis. The table shows the multi-parameter test was statistically significant and the parameter estimates indicate there was a significant EMDA intercept effect. The positive sign of the β_2 parameter estimate indicates that higher EMDA scores were associated with higher initial starting values (i.e., higher intercepts). The positive sign of the β_1 parameter estimate indicate a slight increase in Math Checklist over time.

Discussion

This study demonstrated that short and efficient general outcome measures (GOMs) can work for students with significant cognitive disabilities in math. Number Identification (NI), Number Order (NO), and Number Facts (NF) can work not only practically, but also reliably. These three new GOMs have also demonstrated moderate to strong concurrent and predictive validity with the EMDA and Math Checklist. The results revealed a noteworthy relationship between the new GOMs and the state alternate assessment (MTAS). The MTAS has shown to be a unique criterion measure. As suggested by the correlations produced in this study, unlike the relationship between the GOMs and the two other criterion measures used, only the most difficult GOM, Number Facts, related to the MTAS in a significant way.

Only one of the GOMs, Number Identification, detected student improvement in math across the two years. Importantly, NI was able to detect student growth not only when administered at 3 minutes but also when administered at 1 minute. In addition, the Math Checklist created as part of this study due to the lack of a satisfactory criterion measures for students with significant cognitive disabilities in math detected student growth in math over time. As such, the Math Checklist should be further examined as a potential assessment tool for teachers in addition to the GOMs.

The growth results in this study reflect a frequent observation that students with significant cognitive disabilities learn at slower rates than students with mild disabilities or typically developing students. Consequently, the development of measures for students with significant cognitive disabilities presents itself with a more acute challenge of being more sensitive to smaller changes in performance than is needed for measures developed for higher achieving students.

Even though the results reflect the performance of students at a wide range of grade levels, the reliability, validity and growth coefficients are comparable in robustness to the results on the technical adequacy of GOMs reported in K and 1st-grade general education settings.

This study, even though small in its sample size, demonstrates a great potential for educators teaching math to students with significant cognitive disabilities. If continued to show satisfactory technical characteristics in future research, math GOMs could help teachers plan their instruction more effectively and with more accuracy and sensitivity to student performance and progress in math. In the age of accountability for all students, such performance and progress indicators have the capability to aid these students in succeeding on state tests and in future educational and vocational settings.

It is important to highlight some of the limitations of this study. In addition to the already mentioned small sample size with students ranging across ten grade levels, the GOMs used at different times across the study were of identical forms. In the future, alternate forms of each GOM ought to be implemented to minimize practice effect. Future research should also focus on investigating the technical adequacy of both 3-minute and 1-minute time frames of administration. New measures need to be created to address other math standards, namely measurement and geometry.

References

American Association on Mental Retardation (AAMR; 2002). *Mental retardation: Definition, Classification, and Systems of Supports* (10^{th} ed.). Washington, DC: Author.

Briggs, K. (2005). *Alternate achievement standards for students with the most significant cognitive disabilities: Non-regulatory guidance*. U.S. Department of Education.

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

Browder, D.M., Wallace, T., Snell, M.E., & Kleinert, H. (2005). *The use of progress monitoring with students with significant cognitive disabilities: A white paper prepared for the National Center on Student Progress Monitoring*.

Browder, D.M, Spooner, F., Ahlgrim-Delzell, L., Harris, A.A., & Wakeman, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children, 74*(4), 407-432.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D.J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in students outcomes. *Remedial and Special Education, 29*(1), 46-57.

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

Elliott, S.N. & Roach, A.T. (2007). Alternate assessments of students with significant disabilities: Alternative approaches, common technical challenges. *Applied Measurement in Education, 20*(3), 301-333.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics. *The Journal of Special Education, 41*(2), 121-139.

Fuchs, D., Fuchs, L.S., Benowitz, S., & Barringer, K. (1987). Norm-referenced tests: Are they valid for use with handicapped students? *Exceptional Children, 54*(3), 263-271.

Fuchs, L.S., Hamlett, C.L., & Fuchs, D. (1998). *Monitoring basic skills progress: Basic math computation* (2nd ed.) [computer program manual]. Austin, TX: PRO-ED.

Griffin, S. (2003). The development of math competence in the preschool and early school years. In J.M. Royer (Ed) *Mathematical Cognition*. Information Age Publishing.

Howell, D.C. (2002). *Statistical Methods for Psychology* (5th ed.). Pacific Grove, CA: Duxbury, Thomson Learning.

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

Lembke, E.S. & Foegen, A. (2007). *Indentifying early numeracy indicators for kindergarten and grade 1 students.* Manuscript submitted for publication.

Lembke, E.S., Foegen, A., Whittaker, T.A., & Hampton, D. (2008). Establishing technically adequate measures of progress in early numeracy. *Assessment for Effective Intervention*, 1-10.

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

Martinez, R.S., Missall, K.N., Graney, S.B., Aricak, O.T., & Clarke, B. (2008, November). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention OnlineFirst*.

McConnell, S., McEvoy, M., Carta, J.J., Greenwood, C.R., Kaminski, R., Good, R.H., & Shinn, M. (1998). *Research and development of individual growth and development indicators for children between birth to age eight* (Tech. Rep. #4). Minneapolis, MN: University of Minnesota.

McMaster, K. & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*(2), 85-120.

Minneapolis Public Schools (n.d.). DCD scope and sequence in math. Retrieved August 10, 2006, from http://pic.mpls.k12.mn.us/Scope_and_Sequence.html

Minneapolis Public Schools (n.d.). Minneapolis early childhood special education checklist. Retrieved August 26, 2006, from http://ecfe.mpls.k12.mn.us/

Minnesota Department of Education (2008). *Minnesota Test of Academic Skills (MTAS): An alternate assessment for students with the most significant cognitive disabilities*. Division for Research and Assessment: Roseville, MN.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

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

Perner, D.E. (2007). No Child Left Behind: Issues of assessing students with the most significant cognitive disabilities. *Education and Training in Developmental Disabilities, 42*(3). 243-251.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousands Oaks, CA: Sage Publications.

Research Institute on Progress Monitoring (2007). RIPM Early Numeracy Knowledge and Math Readiness Checklist – Version II. University of Minnesota.

The Psychological Corporation (2002). *Early Math Diagnostic Assessment (EMDA)*. San Antonio, TX: Harcourt Assessment Company.

Wallace, T., Ticha, R., & Gustafson, K. (2009). Technical characteristics of general outcome measures (GOMs) in reading for students with significant cognitive cisabilities. *Reading and Writing Quarterly.*

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

Acknowledgements

Kathy Gustafson

Jeff Long

Appendix

Table 1

Descriptive Statistics: Math GOMs, EMDA and Checklist in Year 1

Measure	Mean	SD	Minimum	Maximum	n
NI 1 min c Winter 07	10.08	5.64	1	19	26
NI 3 min c Winter 07	28.96	18.82	1	**57**	26
NI 5 min c Winter 07	37.64	22.52	1	**57**	25
QD 1 min c Winter 07	11.08	6.92	1	23	26
QD 3 min c Winter 07	31.62	19.68	1	**57**	26
QD 5 min c Winter 07	32.16	20.71	1	**57**	19
NO 1 min c Winter 07	6.42	4.95	**0**	20	26
NO 3 min c Winter 07	14.19	13.30	**0**	49	26
NO 5 min c Winter 07	18.04	17.87	**0**	53	26
Checklist t "yes" Winter 07	29.15	17.05	7	64	26
EMDA t Winter 07	16.35	9.84	2	39	26

Note: c = correct; t = total; NI - Number Identification, QD = Quantity Discrimination, NO = Number Order; EMDA = Early Math Diagnostic Assessment; results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 2

Descriptive Statistics: Math GOMs, Checklist, EMDA, and MTAS in Year 2

Measure	Mean	SD	Minimum	Maximum	n
NI 1 min c Fall 07	11.28	5.25	2	23	25
NI 3 min c Fall 07	31.64	16.77	6	**57**	25
NI 1 min c Spring 08	10.52	4.71	2	18	25
NI 3 min c Spring 08	31.80	15.48	2	55	25
NO 1 min c Fall 07	7.23	4.99	**0**	17	26
NO 3 min c Fall 07	17.23	14.63	**0**	49	26
NO 1 min c Spring 08	8.48	4.78	**0**	17	21
NO 3 min c Spring 08	18.86	15.73	**0**	52	21
NF 1 min c Fall 07	3.24	2.47	**0**	8	25
NF 3 min c Fall 07	7.16	6.14	**0**	20	25
NF 1 min c Spring 08	3.68	3.00	**0**	10	22
NF 3 min c Spring 08	7.45	7.64	**0**	29	22
Checklist t "yes" Fall 07	32.50	17.11	5	70	26
Checklist t "yes" Spring 08	37.19	18.17	7	70	26
EMDA t Winter 08	18.19	11.63	**0**	41	26
MTAS ss Spring 08	196.87	8.73	**172**	214	23

Note. c = correct; t = total, ss = scale score; NI = Number Identification, NO = Number Order, NF = Number Facts, EMDA = Early Math Diagnostic Assessment; MTAS = Minnesota Test of Academic Skills

Results are adjusted for prompting and guessing with a 3 consecutive error rule

Table 3

Math: Concurrent Validity Years 1 and 2

GOM measure	EMDA total Winter 07	Checklist total yes Winter 07	Checklist total yes Fall 07	Checklist total yes Spring 08
NI 1 min c Winter 07	.68**	.73**		
NI 3 min c Winter 07	.64**	.67**		
NI 1 min c Fall 07			.51**	
NI 3 min c Fall 07			.54**	
NI 1 min c Spring 08				.62**
NI 3 min c Spring 08				.62**
QD 1 min c Winter 07	.40*	.39		
QD 3 min c Winter 07	.42*	.44*		
NO 1 min c Winter 07	.65**	.60**		
NO 3 min c Winter 07	.68**	.70**		
NO 1 min c Fall 07			.69**	
NO 3 min c Fall 07			.71**	
NO 1 min c Spring 08				.65**
NO 3 min c Spring 08				.68**
NF 1 min c Fall 07			.56**	
NF 1 min c Fall 07			.59**	
NF 1 min c Spring 08				.78**
NF 3 min c Spring 08				.79**

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

c = correct; GOM = general outcome measure, NI = Number Identification, QD = Quantity Discrimination, NO = Number Order, NF = Number Facts, EMDA = Early Math Diagnostic Assessment
Results are adjusted for guessing with a 3 consecutive error rule

Table 4

Math: Predictive Validity Year 2

GOM measure	EMDA total Winter 08	Checklist total yes Spring 08
NI 1 min c Fall 07	.56**	.50*
NI 3 min c Fall 07	.58**	.52**
NO 1 min c Fall 07	.72**	.70**
NO 3 min c Fall 07	.73**	.69**
NF 1 min c Fall 07	.57**	.41*
NF 3 min c Fall 07	.72**	.58**

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

c = correct; GOM = general outcome measure, NI = Number Identification, NO = Number Order, NF = Number Facts; EMDA = Early Math Diagnostic Assessment

Results are adjusted for guessing with a 3 consecutive error rule

Table 5

Math: Validity with MTAS

GOM measure	MTAS math Spring 08
NI 1 min c Fall 07	.02
NI 3 min c Fall 07	-.09
NI 1 min c Spring 08	-.17
NI 3 min c Spring 08	-.15
NO 1 min c Fall 07	.39
NO 3 min c Fall 07	.36
NO 1 min c Spring 08	.37
NO 3 min c Spring 08	.46
NF 1 min c Fall 07	.20
NF 3 min c Fall 07	.51*
NF 1 min c Spring 08	.53*
NF 3 min c Spring 08	.54*

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

c = correct, t = total; GOM = general outcome measure, NI = Number Identification, NO = Number Order, NF = Number Facts; MTAS = Minnesota Test of Academic Skills

Results are adjusted for guessing with a 3 consecutive error rule

Table 6. *HLM Results for Number Identification 1 Minute (NI1)*

Parameter	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
β_0	4.0137	1.6427	2.4434	0.0073
β_1	1.0286	0.6077	1.6925	0.0453
β_2	0.3842	0.0865	4.4422	< 0.0001
β_3	-0.0493	0.0325	-1.5203	0.0642
$\chi^2(2) = 15.9060$, $p = .0004$				

Table 7. *HLM Results for Number Identification 3 Minutes (NI3)*

Parameter	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
β_0	9.3588	5.6046	1.6699	0.0475
β_1	3.9580	2.0953	1.8890	0.0294
β_2	1.2036	0.2951	4.0782	<0.0001
β_3	-0.1545	0.1118	-1.3824	0.0834
$\chi^2(2) = 14.0316$, $p = .0009$				

Table 8. *HLM results for Number Order 1 Minute (NO1)*

Parameter	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
β_0	0.8561	1.3764	0.6220	0.2670
β_1	0.8628	0.9382	0.9196	0.1789
β_2	0.3423	0.0724	4.7277	<0.0001
β_3	-0.0087	0.0481	-0.1797	0.4287
$\chi^2(2) = 21.5353$, $p < .0001$				

Table 9. *HLM Results for Number Order 3 Minutes (NO3)*

Parameter	**Estimate**	**Std. Error**	***z*- value**	***p*-value**
β_0	-0.9580	3.6962	-0.2592	0.3977
β_1	0.8695	1.8907	0.4599	0.3228

β_2	0.9551	0.1945	4.9093	0.0000
β_3	0.0479	0.0966	0.4958	0.3100
$\chi^2(2) = 19.8881$, $p < .0001$				

Table 5. *HLM Results for the Math Checklist*

Parameter	**Estimate**	**Std. Error**	***z*- value**	***p*-value**
β_0	4.3891	2.9467	1.4895	0.0682
β_1	3.6019	2.0077	1.7941	0.0364
β_2	1.5013	0.1552	9.6714	<0.0001
β_3	0.0255	0.1058	0.2414	0.4046
$\chi^2(2) = 45.8447$, $p < .0001$				

Figure 1. General Outcome Measure – Quantity Discrimination (QD)

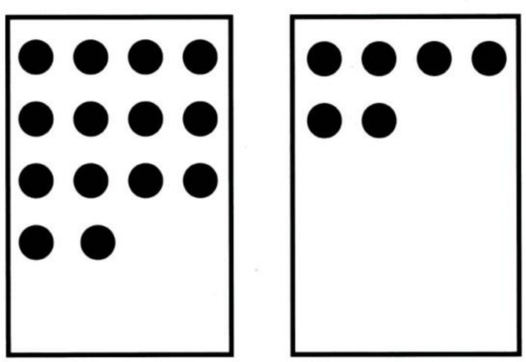

Figure 2. General Outcome Measure – Number Identification (NI)

6 41 52

Figure 3. General Outcome Measure – Number Order (NO)

2	**86**

Figure 4. General Outcome Measure – Number Facts (NF)

1 + 4 = ____

5

4

2

Figure 5. Growth: General Outcome Measures and Criterion Measures

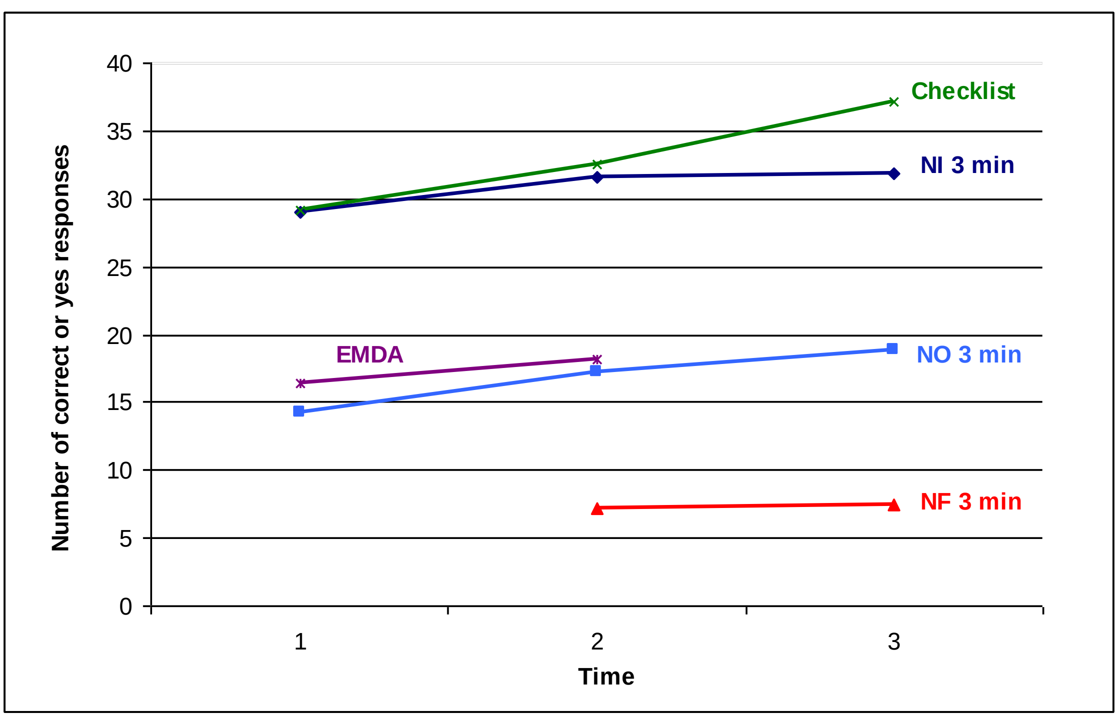

Figure 6. NI1 Individual Growth Curves (thin lines) and Mean Growth Curves (thick line) over Time by Median Split of EDMA (Low/High) and for the Entire Sample (all)

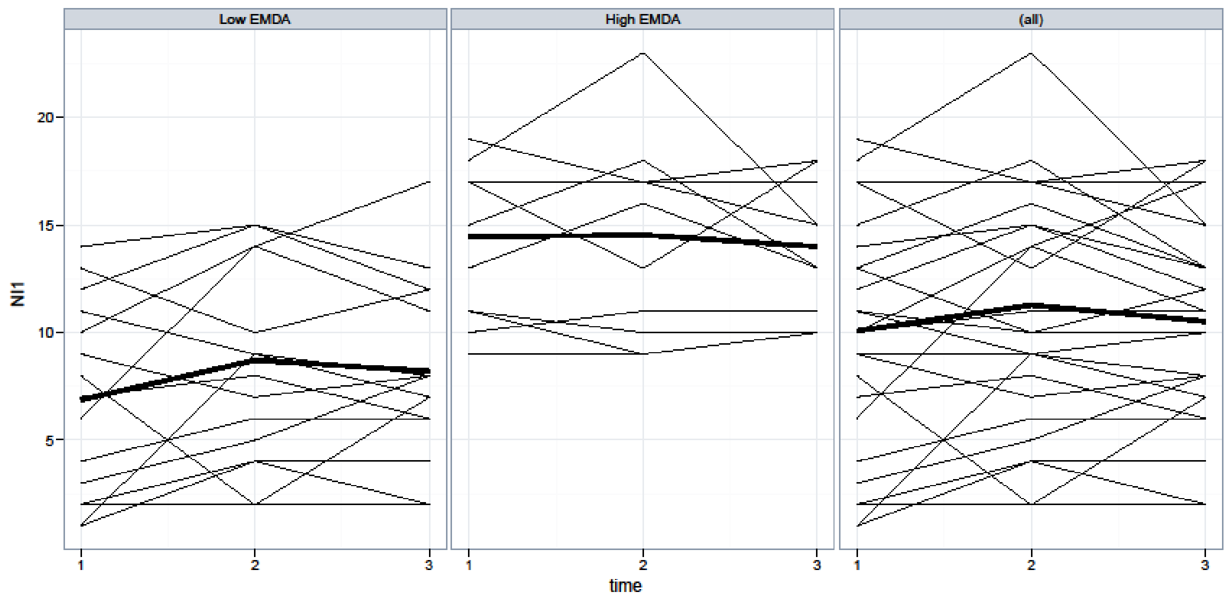

Figure 7. NI3 Individual Growth Curves (thin lines) and Mean Growth Curves (thick line) over Time by Median Split of EDMA (Low/High) and for the Entire Sample (all)

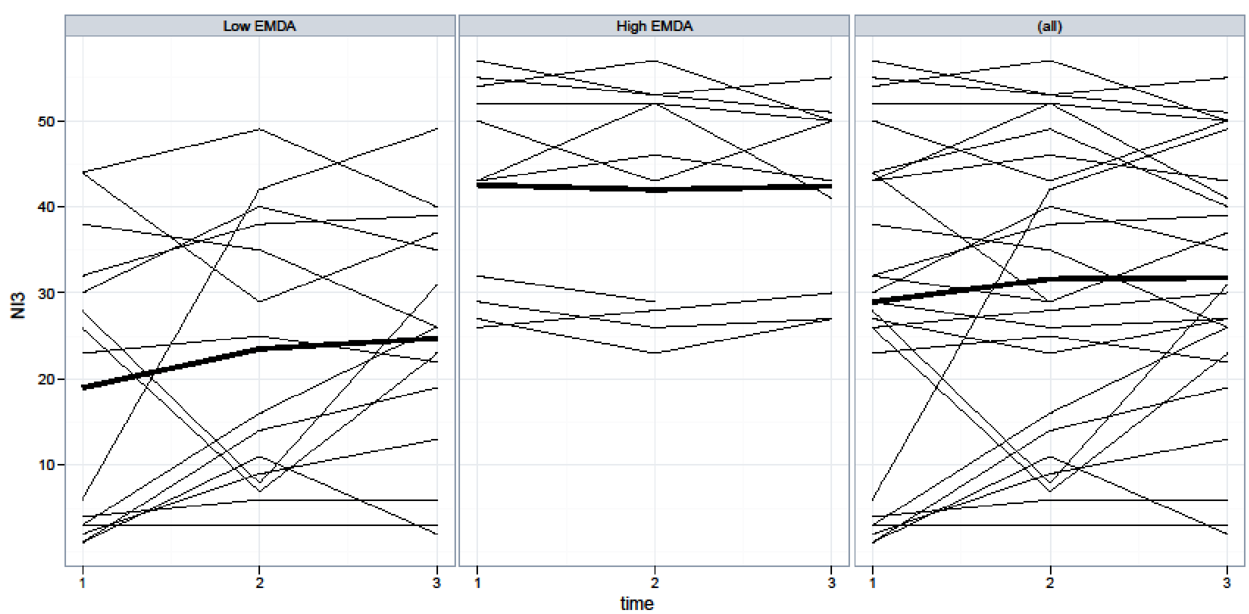

Figure 8. NO1 Individual Growth Curves (thin lines) and Mean Growth Curves (thick line) over Time by Median Split of EDMA (Low/High) and for the Entire Sample (all)

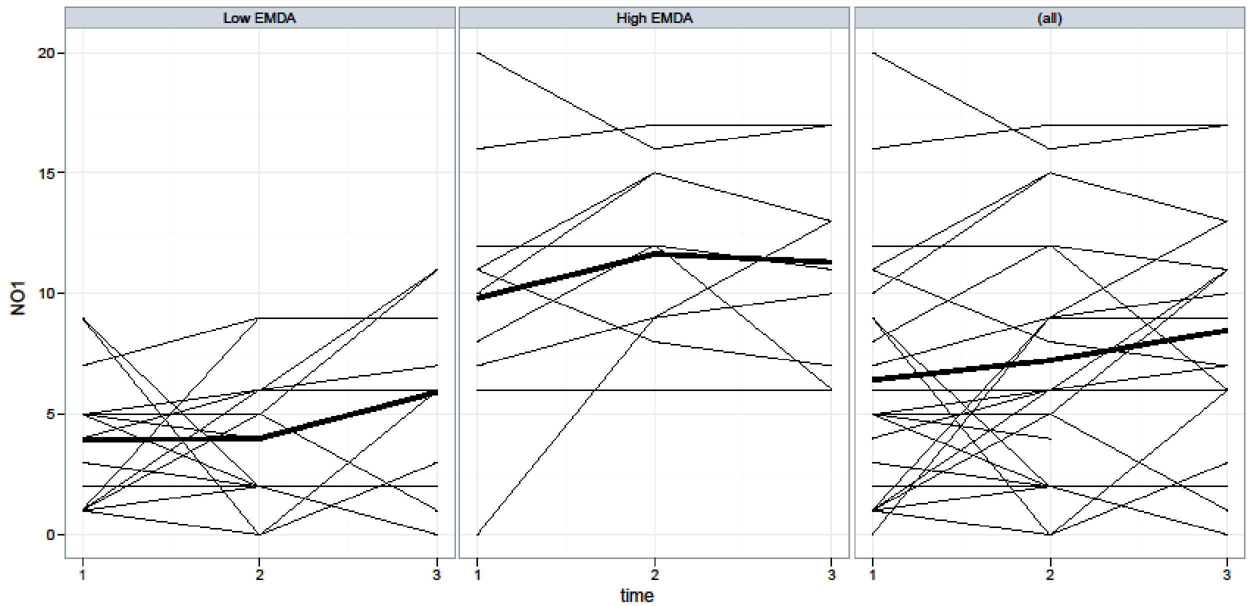

Figure 9. NO3 Individual Growth Curves (thin lines) and Mean Growth Curves (thick line) over Time by Median Split of EDMA (Low/High) and for the Entire Sample (all)

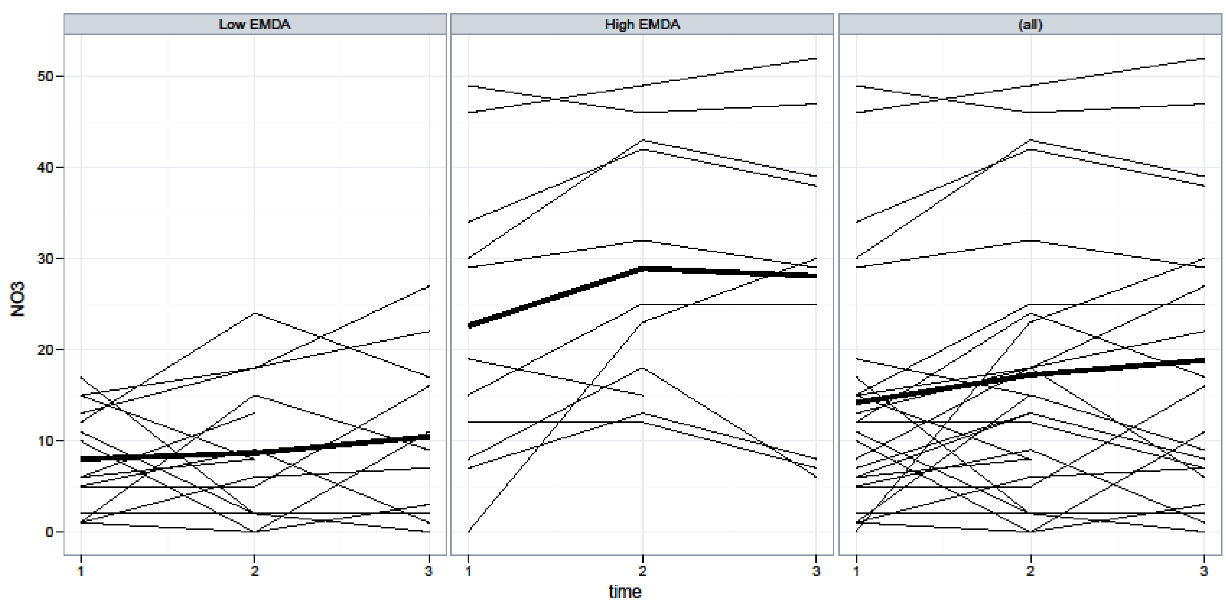

Figure 5. MC Individual Growth Curves (thin lines) and Mean Growth Curves (thick line) over Time by Median Split of EDMA (Low/High) and for the Entire Sample (all)