RELIABILITY, VALIDITY, AND EQUIVALENCY OF A COMPUTERIZED CURRICULUM BASED MEASURE FOR BASIC MATH SKILLS

By

CASSANDRA BRENNER WONG

Bachelor of Science in Psychology
Cameron University
Lawton, Oklahoma
2004

Master of Science in Educational Psychology
Oklahoma State University
Stillwater, Oklahoma
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July 2009

RELIABILITY, VALIDITY, AND EQUIVALENCY OF

A COMPUTERIZED CURRICULUM BASED

MEASURE FOR BASIC MATH SKILLS

Dissertation Approved:

Dr. Terry Stinnett

Dissertation Adviser

Dr. Gary Duhon

Dr. Robert Davis

Dr. Dale Fuqua

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to extend a special thanks to all who made this project possible. First, I could not have made it through this research project or graduate school without the help and support of my husband, Victor.

I also greatly appreciate the help given to me from my advisor and director of training in guiding my work as well as stepping out of the way so that I may learn for myself.

My research team on this project was invaluable and could not have been more willing to help and learn what was necessary for this project.

The schools that allowed me to come in and take over their computer lab for a couple of days and call many of their students out of class did me a great favor.

Finally, I would like to thank my children, Kaleb and Addison, for their patience and understanding when mommy had to work all the time and could not play with them. I also want to thank them for keeping me grounded while I was going through this whole process.

.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Curriculum Based Measurement (CBM) is an assessment method used by school based personnel designed to provide information about student functioning within the current curriculum (Shapiro, 2004). These measures provide direct assessment of academic performance and have been shown to be reliable, repeatable, sensitive to student growth, and assistive in determining appropriate instructional strategies (Shapiro, 2004). Derived from student's current curriculum CBM provides information directly related to the student and reduce the problem of poor overlap between the curriculum and the test that is often associated with standardized achievement test (Shapiro, 2004). Research supports the use of CBM for screening decisions, eligibility decisions, progress monitoring, and program evaluation (VanDerHeyden, A., Witt, J. & Gilbertson, 2004). The reliability and validity of CBM has been established through multiple studies (Allinder, Fuchs, & Fuchs, 1998). In a review of CBM computation reliability studies Marston (1989) identified internal consistency for mixed probes of .93; test-retest was reported at .93 for mixed probes; parallel form reliability ranged from .48 to .72; and inter-scorer reliability was found to be .93 to .98 on mixed probes. CBM math initially provided lower correlations than other academic areas. Criterion validity correlations were .27 to .67 between computation probes and district criterion referenced tests (Marston, 1989). More recently, criterion validity of CBM math procedures was found to be .66 to .91 with problems correct and .77 to .87 with digits correct when compared to the Math Computation Test – Revised (Fuchs, Fuchs, Hamlett & Stecker, 1991). A study conducted in 2006 also reported higher reliability and validity when examining the relationship of CBM (computation, concepts/applications, reading) to standardized assessments: Pennsylvania System of School Achievement (PSSA), SAT 9, MAT-8 and Stanford Diagnostic Reading Test (SDRT; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006).The reliability and validity of CBM has been evaluated in numerous studies and has repeatedly been shown to be at the moderate to very strong level(Marston, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991; Deno, 1985).

Although CBA has numerous benefits and is one of the most common forms of assessment in the classroom, a number of concerns have been identified which may negatively impact utilization and interpretation of CBA. A frequently mentioned concern with the use of CBA is the potential problem of scoring errors (Fuchs, Fuchs, & Hamlett, 1994). These measures are often scored by the teacher or teacher's aid which may lead to human error in scoring as well as data entry. Another concern is the time and cost associated with developing, printing, administering, scoring, and reporting of CBA results. Further teachers also find it difficult to convert systematic assessment information into meaningful instructional changes that can be implemented feasibly within their educational programs which typically serve the entire class. It can be difficult for a teacher to make an instructional change to review subtraction with regrouping when the majority of the class has learned the skill while a lesser number is struggling. This is especially true when the teacher may be trying to finish a program of curriculum within a certain time frame. These types of problems are the most often heard complaints about CBM (Deno, 2003). A study conducted by Fuchs and colleagues addressed the time commitment complaint with the solution of using computers to develop, administer, score, graph and interpret results. This study found an average of seven minutes time savings by using the computer (Fuchs, Fuchs, Hamlett, 1992). Although the measures are normally brief, these problems associated with the typical paper administration supports the necessity of modification in attempt to alleviate some of these problems. The impact of these problems could be detrimental for students and teachers and the cost of developing and printing the probes may make the use of CBM aversive for the school administrators because of budgeting restrictions. Inaccurate scoring may lead to instructional changes that are not necessary, the time commitment may adversely affect the time the teacher has for instructional time, or the problems could lead to reluctance to use CBM at all, despite it's benefits.

In the search for an alternative to the limitations of the paper mode, computer technology is a frequently used solution (Mead & Drasgow, 1993). Although traditional CBM is widely used in spite of the above mentioned difficulties, these concerns could be reduced or even eliminated by the use of computerized versions of CBM. For years computerized math programs have been used in spite of the lack of empirical support for this mode of administration. Since time commitment required by the teacher is one of the top complaints, administering the math probes via computer reduces the time consuming tasks of scoring and entering data to do comparisons as well as reducing the possibility of human error. By using a computer to administer, it is also possible to calculate results immediately, graph class performance, as well as make intra-individual comparisons. This work can all be performed by the computer and takes little or no teacher time to complete. The use of computer eliminates the problem of human error in scoring and provides a uniform scoring system. With any assessment it is necessary to have valid results, especially if these results are going to be used to make important educational decisions. The use of computers provides a fast, accurate and inexpensive way to administer, score, and report CBA results, however before computerized versions of CBM can be utilized with confidence the reliability and validity must be established. Alternatively, studies of equivalence between two versions of a test are needed to allow the continued use of normative and validity information developed from the paper and pencil version. This evidence should include similar score distributions and rank orders of scores for the same individuals from both versions of the test (Pomplun, Frey, & Becker, 2002).

Although there is a lack of research supporting the reliability, validity and equivalency of computer CBM relative to paper versions, there has been considerable literature in the area of test format equivalency for other assessment types. Test equivalency refers to when there is no significant difference between test scores (Pomplun, Frey, & Becker, 2002). In a review of educational assessments comparing performance on computer forms of assessment to currently established paper versions, three studies showed higher scores on computer forms, eleven studies found no differences, and nine studies showed higher scores for paper based tests. Most studies find that the differences in scores between paper and computer forms are generally small

(Clariana & Wallace, 2002; DeAngelis, 2000; Hargreaves, Shorrocks-Taylor, Swinnerton, Tait, & Threlfall, 2004).

In order to use the alternative method of computer based CBM it is necessary to establish that it can reliably and validly sample student performance. Although research has provided support for adequate validity and reliability of Math-CBM (Fuchs, Fuchs, and Courey, 2005) when administered in paper form, no research to date has established the reliability and validity of a computer based form of CBM. In a series of studies the authors have attempted to begin the search for these answers. The first study sought to establish the reliability and validity of a computerized CBA/M program by comparing it directly to its supported paper equivalent. Both the computerized and paper versions were found to be reliable across multiple administrations. Further a .80 correlation was found between the means of the paper test and the means of the computer test. The paper scores were found to be consistently higher (Duhon, Wong, & Mesmer, 2007).

This study seeks to determine if a computerized CBM procedure as reliable, valid and equivalent. Further three modes of administration will be compared and differences will be evaluated using a Multivariate Analysis of Variance. It is hypothesized that computerized CBM will be a reliable measure. It is hypothesized that there will be differences between the three modes of testing. It is also hypothesized that the relationship between paper and computer CBM will be increased as visual similarity between formats increase. Finally it is hypothesized that computer performance will effected by the grade level which a participant is currently in.

CHAPTER II

REVIEW OF LITERATURE

Although in recent years there has been movement in the teaching of mathematics towards emphasizing problem solving skills rather than computational competence, it is reasonably argued that a foundation in computational skills leads to success in all areas of mathematics (Shapiro, 2004). The field of school psychology develops and produces practitioners with the skills to fulfill a variety of roles. School psychologists are trained to administer a wide range of psychological assessment, intervention, prevention, health promotion, program development and program evaluation. It is their goal to "promote positive learning environments within which children and youth from diverse backgrounds have equal access to effective educational and psychological services to promote healthy development" (Fagan & Wise, 2000). School psychologists have specialized training and experience in both education and psychology. As outlined recently by Layton and Lock, Curriculum based assessment is a assessment procedure that fulfills partial requirement as set may the law No Child Left Behind which corresponds closely with the new Individuals with Disabilities Education Improvement Act (IDEA; 2007).

This review of the literature will include the history of curriculum based assessments, the technical adequacy of these instruments and the limitations of these instruments.

Common uses of the instruments will be explored and well as recent developments and new uses. Further, computers as the solution to the limitations of CBA/CBM will be covered with considerable evidence of this solution being successful for other instruments. Finally the logical path leading to the current study will be explained while not excluding information about possible beneficiaries of the findings.

There are three terms that are prevalent in the education literature with regard to curriculum measurement: Curriculum Based Assessment, Curriculum Based Measurement, and Curriculum Based Evaluation. Curriculum Based Assessment (CBA) is a broad term used to describe any and all assessment activities that are constructed using a student's instruction materials. It was developed in the 1970's to provide teachers with a quick, accurate, and simple way of evaluating if their students had learned what was taught. This approach emphasizes direct, repeated measurement of a target academic area. Probes for each academic area are developed and used to collect data on the student's performance (Witt, Elliot, Daly, Gresham, & Kramer, 1998). Multiple forms of these tests may be created and then administered to evaluate fluency or proficiency and accuracy. A mastery criteria is set using the average or typical level of a median performing student is the class and the agreement of an assessment team (teacher and special education teacher, typically). Criteria that are nationally normed or locally normed may also be used. Curriculum assessments can be and have been used as part of multidisciplinary assessments, Responsiveness to intervention (RTI), development of local norms, and for classroom use by teachers. Curriculum-Based Assessments can be used as an effective, efficient and accurate model of special education decision making. It is used in five main areas of decision making: screening and referring, determining eligibility and classifying, planning programs, monitoring progress and modifying programs, and evaluating programs (Cundari & Suppa, 1988). A variety of administrators use curriculum assessments including teachers, specialty teachers, school psychologists, and other school personnel.

Curriculum Based Measurement (CBM) is a more precise form of CBA. It is a method designed for teachers to use in monitoring student progress and to assist teachers in improving student's academic performance. CBM consists of a standardized format and procedures and provides a technically strong assessment system for school psychologists, special education teachers and administrators (Allinder, Fuchs, and Fuchs, 1998b). CBM originated as part of the Data-based Program Modification model. This model targeted the students in special education by detailing a process of using progress monitoring data to make educational programming decisions. It was designed to help special education and resource teachers improve their academic interventions. In order to evaluate whether this model measures what it claims to measure a federally funded study was conducted through the University of Minnesota Institute for Research on Learning Disabilities (IRLD; Deno & Mirkin, 1977). The research demonstrates that teachers were more effective in developing interventions when using such a model (Fuchs, Fuchs, & Hamlett, 1989). Some common uses for CBM have been screening, referral, and identification for special education services (Shinn, 1989), program evaluation, and reintegration of students with disabilities into regular education classrooms (Allinder, et al, 1998b;Tindal, 1992). In CBM, the teacher or school psychologist uses standardized administration and scoring procedures in which the student is presented with short, timed probes. The probes used are typically single skill probes. Therefore for each academic area and within that academic area each separate skill would be assessed individually. The fairly agreed upon distinction between CBA and CBM is that a CBA is often a collection of all the skills a student should learn within a grade and their performance on the measure tells the teacher what areas the student needs to improve on. CBM, on the other hand are measures of a single skill or skill area such as addition without regrouping or subtraction with regrouping to tell the teacher the level at which the student is performing that specific skill.

Curriculum Based Evaluations (CBE) as defined by Deno in 1985 is any set of measurement procedures used to gather information that may be used for educational decisions. These procedures use direct observation and a student’s performance in the local curriculum. This model of Curriculum-based Assessments as proposed by Howell and Morehead (1987) suggests that students be tested over the subcomponents of instructional tasks. The results assist in the design of the intervention by identifying skills that are missing from the student’s repertoire (Howell & Morehead, 1987). It should now be obvious that these three distinctions of curriculum assessment are very similar and are often only distinct in the way the results are used or presented. There is considerable research in the area of curriculum assessments and a review of the reliability and validity is necessary when proposing the use of any assessment procedure.

Reliability of CBA/CBM

In general, CBM has been found to be a psychometrically strong procedure, making the reliability and validity of CBM a distinct advantage for school psychologists. CBM’s technical adequacy has been established through multiple studies across several years (Fuchs, Deno, & Marston, 1983; Marston, 1989; Tindal & Parker, 1991). While there are many studies to support the technical adequacy of CBM, the studies that have investigated CBM-math have found lower but still adequate and consistent reliabilities and validity (Marston, 1989). Math probes may include single type problems or multiple type problems. Specifically, the problems may also be computation or application problems. Test reliability is when a test's observed scores correlate highly with its true scores (Allen & Yen, 1979). Three common methods are used to estimate the reliability coefficient: test/retest, alternate forms and internal consistency. Test-retest reliability refers to the stability of test scores over time. Therefore the degree to which a test given at one time correlates with the same test given after a period of time has passed. Alternate forms reliability refers to the extent to which two different but equal forms of the same test correlate with each other. Reliability can be expressed as the correlation coefficient between the observed scores of parallel tests. Another term for parallel tests is also alternate forms test reliability. Alternative test forms are two tests that have be developed, making the effort for them to be parallel, that typically have similar score means, variances, and correlations with other measures. The correlation between observed scores on the alternate forms estimates the reliability of each form as well as the how parallel the tests are. This is obviously one type of reliability that this project focuses on (i.e. the reliability of a paper test compared to the alternate form computer test). Finally, internal consistency addresses the degree of homogeneity of the items of a test. In order to have high internal consistency all the items of a test would be measuring the same construct (Kazdin, 2003). In the area of CBM Reading, technical adequacy has been well established (Deno, 1985, Shinn, 1989, 1998); and more recently the technical adequacy of CBM math as well. Some studies conducted, have found high inter-rater agreement

(.97), high 1 week test-retest reliability (.87), and moderate alternate form reliability (.66) have been demonstrated for CBM-Math (Tindal, Marston, & Deno, 1983). Another study found high alternate form reliability (.91) but somewhat lower inter-scorer agreement (.83) (Thurber, Shinn, & Smolkowski, 2002). Five types of reliability are reported in a summary of computation studies. Internal consistency was reported as .93 for mixed probes; test-retest ranged from .78 for division-only probes to .93 for mixed probes; parallel-form (as known as alternate forms) reliability ranged from .48 to .72; and inter-scorer reliability ranged from .93 to .98 on mixed probes (Marston, 1989).

Validity of CBA/CBM

While it is necessary that a test reliably measures what it is intended to measure, it is also essential that a test be found to show validity. This section describes the concept of validity and discusses the validity of CBA and CBM. Validity is the degree to which a test measures what it claims to measure. There are three major types of validity: content validity, criterion-related validity, and construct validity. Content validity is the extent to which a test measures the content of the intended area. Content validity can be subdivided into item validity and sampling validity. Item validity deals with the degree to which a test's items are relevant to the intended area. Sampling validity deals with the degree to which a test measures the total content of the targeted area. Content validity is established through a rational analysis of the content of a test and does not require statistical calculations (Allen & Yen, 1979). For example in the case of a math CBA for third grade math skills, the test would have content validity if it sampled all of the skills learned in third grade. It is determined by expert judgment using a process of looking at the proposed test and comparing what is taught and what is tested. When these two things coincide, there is high content validity (Gay & Airasian, 2003).

The second major type of validity is criterion-related validity which is demonstrated when the test score produced is related to some criterion. An example of this would be when the results of the CBM correlate highly with teacher assigned grades. In this case the behavior that the test scores are used to predict (i.e. teacher assigned grades) is the criterion. Criterion-related validity may be categorized as predictive or concurrent validity. Predictive validity refers to the correlation between one measure and the future performance on another criterion or measure. Concurrent validity involves examining the scores on a test and comparing them to another established test or criterion. To establish concurrent validity it is important that both tests be given within the same time frame (Gay & Airasian, 2003). Concurrent validity is the type that most closely describes the purpose of this research study since the literature has yet to establish computerized curriculum measures as a valid alternative form of measurement for paper and pencil CBM.

Curriculum based assessment and curriculum based measurement procedures are designed to show how a student is performing currently. By using the measures repeatedly it is used to show growth. When comparing CBM to commercial math tests, there were correlations exceeding .60 while a moderate correlation of .43 was found for the Problem-Solving subtest and .54 with the Math Operations subtest of the Metropolitan Achievement Tests (MAT). Examining the validity of CBM-math, a study was conducted using computerized CBM-math procedures and Math Computation Test-Revised (MCT-R) and reported criterion validity at .66 to .91 for problems correct and

.77 to .87 for digits correct. Correlations were also found for the computerized CBM-math and concepts of Number (NC) from and Math Computation (MC)subtests of the Stanford Achievement test. This investigation reported correlations of .49 to .88 for NC and .55 to .93 for MC (Fuchs, Fuchs, Hamlett, & Stecker, 1991). These studies produce as compared to another criterion measure. This type of validity is vital to evaluating the usefulness of a new assessment. However, the third type of validity is just as necessary for this purpose.

CBM has also been shown to have good construct validity which is the degree to which a test measures the theoretical construct it was supposed to measure (Allen & Yen, 1979). A study by Shinn and Marston (1985) demonstrated construct validity for CBM math procedures. They found that students could be differentiated among regular education, Title I, and Special education at grades 5 and 6. By distinguishing between these groups of students with differing characteristics, construct validity was established in this case. There are two types of construct validity: convergent and divergent validity. Convergent validity is when a test correlates well with other tests believed to measure the same construct. Divergent validity is just the opposite. For a test to have divergent validity it should not correlate well with other tests that are believed to measure a unrelated construct. A study focusing on construct validity found that CBM-Math correlated highly with measures of basic math facts (r =.82) and moderately with measures of math computation (r =.61). However it correlated less well with measures of math applications (r =.42; Thurber, Shinn, & Smolkowski, 2002).

CBA/M has established reliability and validity through multiple studies (Allinder, Fuchs, & Fuchs, 1998). CBM math initially provided lower correlations than other academic areas. Criterion validity correlations were .27 to .67 between computation probes and district criterion referenced tests (Marston, 1989). More recently, criterion validity of CBM math procedures were found to be .66 to .91 with problems correct and .77 to .87 with digits correct when compared to the Math Computation Test – Resised (Fuchs, Fuchs, Hamlett & Stecker, 1991). A study conducted in 2006 examined the relationship of CBM (computation, concepts/applications, reading) to standardized assessments: Pennsylvania System of School Achievement (PSSA), SAT 9 , MAT-8 and Stanford Diagnostic Reading Test (SDRT; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). Testing was conducted three times a year: Fall, Winter, and Spring. The correlations between CBM-math computation and PSSA ranged from .51 to .53 for Winter and Spring time periods with all correlations being significant at $p<.001$. Correlations between CBM-math concepts/applications and PSSA ranged from .56 to .64 for Winter and Spring, with all correlations being significant. The SAT-9 showed very strong correlations with CBM – computation and concepts/applications for Winter and Spring again (Shapiro et al., 2006). These findings are consistent with another study which compared CBM to state assessments for students with a learning disability (Helwig, Anderson, &Tindal, 2002). Using discriminant analysis Helwig et al. (2002) was able to predict with 87% accuracy which students would meet the state math standards through the use of CBM math probes. Further, Shapiro et al. (2006) examined the predictive power of CBM math and reading on PSSA and found correct classification rates of 66% to 85%. These studies support the use of CBM probes to predict performance on state assessments and to monitor classroom progress both of which are important aspects of instructional planning (Shapiro et al., 2006).

Use of CBA/CBM procedures

Over the last decade CBA procedures have received a considerable amount of attention in the empirical literature as a reliable and valid measure of academic skills. Currently there are two studies that address the use of, daily practice of and attitudes toward CBA use in professional practice. Members of the National association for School Psychologists (NASP) were participants for a 1990 study in this area. Researchers reported that 46% of respondents used CBA procedures in daily practice with 18% reporting high and consistent use. In a follow up study using the 1999-2000 membership of NASP, researchers conducted a similar analysis. It was found in the follow up study that 54% of respondents use CBA procedures on a daily basis and now 32% indicated frequent use (Shapiro, Angello, & Eckert, 2004). This data demonstrates a moderate increase. However it should be noted that CBAs are still consistently in use and the frequency of their use by professionals is increasing.

CBA is used to answer certain questions with regard to a child's performance such as what does the child know, what can he do, how does he think, what is his approach to a task that is unknown, and finally what should the teacher do now. It is important to know the child's ability level to help him learn at a level that is most beneficial for learning. A child performs a task that is typically drawn from the curriculum and his performance is compared to that of peers his same age.

Although CBA/CBM has a long history in the literature there have been some recent advances to make the use of CBA/CBM procedures more beneficial to teachers and school psychologists. Technology has been employed to provide teachers with graphs of progress and a skills analysis to guide instructional planning. A criticism of CBA is that while it identifies the skill deficits, it does not provide interventions to address the problem (Shapiro et al., 2006). Essentially it tells what is wrong but not how to fix the problem. Monitoring Basic Skills Progress (MBSP) developed by Fuchs, Hamlett, and Fuchs is software version of CBM. This program was normed on 12,000 participants, including a 5% population for children classified as having a learning disability. For MBSP math, the student takes the CBM probes via the computer weekly which provides the teacher with two scores: an overall competence of the skills assessed and individual skills mastery scores. The program also automatically graphs student progress and analyzes that progress with regard to the goal as set forth by the teacher. This type of skills analysis has shown strong treatment validity. Experimental studies examining treatment and control groups have found significant effects when using the skills anaysis and expert systems in reading (Fuchs, Fuchs, Hamlett, & Ferguson, 1992), spelling (Fuchs, Fuchs, Hamlett, & Allinder, 1991, and math (Fuchs, Fuchs, Hamlett, and Stecker, 1991). Teachers that use these systems are found to use a diverse set of teaching strategies, have the ability to reteach both wide range and specific components of instruction, and attempt to increase motivation in students (Fuchs, Fuchs, & Hamlett, 1994). The reliability and validity of the skills analysis was investigating producing reliability for math of .83 (problem type by problem type agreement at different mastery levels with one week intervals) and a validity correlation of .73 (skill by skill analysis correlated with the median scored graphed; Fuchs, Fuchs, Hamlett, & Allinder, 1989).

Limitations of paper CBA/benefits of computer

Although CBA has numerous benefits and is one of the most common forms of assessment in the classroom, a number of concerns have been identified which may negatively impact utilization and interpretation of CBA. One of the most frequently mentioned concerns with the use of CBA is the potential problem of scoring errors (Fuchs, Fuchs, & Hamlett, 1994). CBA measures are often scored by the teacher or teacher's aid which may lead to human error in scoring as well as data entry. An equally concerning issue with CBA's is the time and cost associated with developing, printing, administering, scoring, and reporting of CBA results. These types of problems are the most often reported complaints about CBA/CBM (Deno, 2003). Although the measures are normally brief, the teacher must develop a probe, multiple alternate forms for repetition, print the probes, administer them, score them, and enter the results into a graphing program or a program designed to give a skills analysis (Fuchs, Fuchs and Courey, 2005). These tasks can consume a considerable amount of teacher or administrator time. When using traditional psychological tests to assess children in the schools, there are personnel assigned almost exclusively to those tasks (i.e. school psychologists, educational diagnosticians). Therefore, one limitation to the utilization of CBA is that it makes the teacher responsible for student assessment, a change that may not be well received by teachers. CBAs also require time and a considerable amount of work to compare students within the class and to other classes. It is for these reasons that teachers who already have significant demands on their time may not feel able or may feel overwhelmed by the added responsibility. Other areas of assessment have tackled similar issues and found computer administration as a viable solution. It has previously been demonstrated that the use of technology to administer probes can greatly reduce teacher time (Fuchs, Fuchs, & Hamlett, 1992).

Computer/paper equivalency

With the consideration of the limitations associated with paper CBA/CBM, it is necessary to investigate other more efficient options. An assessment can take many forms: paper, oral, or computerized. The use of computers in assessment has grown rapidly during the 1980s and 1990s while the use of computers in the schools has also increased rapidly. In 1992, 3.5 million computers were found in the schools (Allinder, Fuchs, & Fuchs, 1998b). However, this increased to 8.6 million reported in 1998, which is an increase of over 150%. Since then, the number of instructional computers in schools has grown by approximately 15% a year.

There are a number of explanations for this increased growth of computer use in schools and for psychological assessment. The increase in caseloads for human service providers coupled with budget cuts for social programs certainly played a role. Of 268 school psychologists surveyed, 72 % reported they used computers in their assessment procedures and even more reported the intent to use computers in the future (Allinder, Fuchs, & Fuchs, 1998a). After a through review, Mead and Drasgow (1993) reported that clinical instruments, personality scales, job attitude surveys, and cognitive tests are just a few of the psychological tests that have been computerized. The transfer to computer assessment provides the following advantages: ease of administration, reduced training requirements, faster scoring, fewer scoring errors, and fewer opportunities for cheating. Although reliability has already been briefly discussed it is necessary here to discuss “alternate-forms reliability” because the topic is key to any discussion regarding computerized forms of testing. While literature has not addressed the equivalence of paper and computer CBA/CBM measures, the equivalence of these two modes of administration have been evaluated for other tests. Test equivalency refers to when there is no significant difference between test scores (Pomplun, Frey, & Becker, 2002). For example, a meta-analysis of the paper and pencil versus the computerized GRE found that the verbal and quantitative abilities measured by both types of assessment are the same. When developing the Hamilton Anxiety Scale, it was found that the computerized format correlated very high (r = .92) with the traditional paper format (Allinder et al., 1998a). In yet another comparison, the computerized and paper formats of the Myers-Briggs Type Indicator were found to have high correlations (r = .84 to .86; Allinder et al, 1998a). Similar to the studies just mentioned, there have been a number of other studies comparing paper and pencil forms to computer forms of various tests. Curriculum Based Measurements for math and writing are often performed using paper/pencil format while reading CBMs may take the form of oral or paper/pencil. There is also recent literature regarding CBM facilitated by the computer. In 1990, Fuchs, Hamlett, and Fuchs presented computer software designed to facilitate CBM probes called Monitoring Basic Skills Progress which includes Basic Math, Basic Spelling, and Basic Reading. This program is designed to present information to alert the teacher of problem areas in which changes in instruction are needed and to inform the teacher of the need to increase goals for students performing above goal. For math and spelling the program also graphs the overall growth and acquisition of each skill for each child. Computerized CBM testing for math normally consists of a computerized presentation of mixed math problems from the year's curriculum. Spelling requires that a person dictates words while the students type them in to the computer. Reading entails the students reading a 400 word passage on the computer that has every seventh word deleted and replaced with a drop down multiple choice or performance with oral reading fluency. In oral reading fluency on the computer the teacher may click on the missed words as a child reads and click on the last word ead when the time has elapsed. All probes use short time limits that vary by skill or grade ranging from 2 to 5 minutes. The results can then be viewed by the teacher and graphs are developed that compare the actual progress to the desired goal progress.

Support was found for using computerized CBM with skills analysis by using three groups of teachers: CBM with graphed analysis, CBM with graphed plus skills analyses, and no CBM. This type of study was conducted for both spelling and math. It was found that teachers using computerized skills analysis were better able to adapt their curriculum to the students needs and produced higher achievement than the CBM with graphed analysis and the no CBM groups (Fuchs, Fuchs, Hamlett, & Stecker, 1991; Fuchs, Fuchs, Hamlett, & Allinder, 1991). The above study also addresses some important criticisms that have been found with regard to curriculum based testing. As mentioned, problems are the time consumed by administering, scoring and interpreting the testing. The time commitment of the teacher using CBM is one of the top teacher complaints. Another study directly addresses the amount of time saved by using computer administration. The study randomly assigned teaches to two groups: one that collected CBM data in all areas and scored them by hand and the other administered and scored the measures by the computer. The researchers observed a time savings of 7 minutes per student by using the computer rather than the traditional method. When carrying even a small case load of twenty students that are measured twice weekly, a teacher can save up to 5 hours a week in administration and scoring duties by using the computer (Fuchs, Hamlett, Fuchs, Stecker, and Ferguson, 1988). This is a considerable amount of time savings.

There have been studies to compare paper and pencil and computer forms of various other tests and there have been computer programs designed to administer, score and interpret curriculum based measures. However, there is a lack of research that directly compares paper and pencil forms to computer forms of curriculum based assessments for math fluency.

Paper-based vs. Computer-based Administration

There are a number of things to consider when attempting to develop and evaluate a new form of test such as technical adequacy, item construction, and form decisions. The new form of interest here is the computerized form of curriculum based assessment of math fluency. Research indicates that there have been an array of findings when comparing paper and pencil forms of tests to their computerized equivalent. For example, for Mathematics and English CLEP tests, the paper based scores were higher than the computer based scores. The same was found in a study of fighter plane silhouettes while DeAngelis (2000) found higher scores on computer based tests than paper based for a dental hygiene course unit midterm examination (Clariana &Wallace, 2002). However, other similar studies found no significant differences. In a review of educational assessments, three studies showed higher scores on computer forms, eleven studies found no differences, and nine studies showed higher scores for paper based tests. Most studies find that the differences in scores between paper and computer forms are generally small. However, the effects of these differences may be large for a particular examinee.

Individual differences studied include content familiarity, computer familiarity, and non-competitiveness (Clariana & Wallace, 2002). Other individual differences evaluated are gender, race and age. In a 1993 study, white males performed better on computer delivery of the Graduate Record Examination (GRE) while males from other racial groups excelled on the paper format. There were no differences found for females. In this particular study there were also no differences found for format preference or computer familiarity. As well, a study combined factors of individual differences: content familiarity, computer familiarity, competitiveness, and gender in the evaluation of paper and computer modes of a general computer knowledge exam. Researchers found that the only factor that differed was content familiarity (Clariana & Wallace, 2002). There has been support for the completion time differences between paper and computer administered tests. Some researchers found the computer tests were completed faster while others found that they took longer for people that demonstrated minimal computer skills and finally computer tests were found to take longer on the first administration but not on subsequent times (Bodmann & Robinson, 2004). Perhaps the most supported explanation for test modes differences is the lack of flexibility that most computer based test provide. Studies have been conducted to examine score and speed differences between modes. Bodmann and Robinson used random assignment with undergraduates that were familiar the interface to investigate score differences and speed differences of a test of educational psychology. Tests covered content in the educational psychology course and contributed to the students grades. The Computer based test used ActiveInk and presented items one at a time in a multiple choice format. The CBT operated with the following characteristics: items can not be reviewed or changed; items can not be skipped or returned to later. Half of the students took PPT and the other half took the CBT for the first test, this was reversed for the second test. Both tests were timed for each individual student. For the test scores there were no difference between the PPT and the CBT, t(53)

= .04, p = .97. The test times did show a difference of four minutes, the PPT being longer, t(53) =3.95, p <.001. A follow up experiment was conducted to examine the test time differences with the hypothesis that the flexibility of the PPT increased completion time. A different educational psychology class served as the participants to this experiment. A total of three test were given, using these three modes: “all at once, scroll”(all items presented at once, most similar to PPT); “one at time, revisit”(items presented one at a time, but could revisit and change answers before submitting test); and one at time, no revisit”(same as experiment one). All students received each test mode and the order was counterbalanced. Repeated measures ANOVAs were conducted showing no differences in test scores, F(2,114)=0.77, MSE = 6.76, p=.46. Differences were found for test times, F(2,114) = 12.24, MSE = 1346.5, p<.001. Post hoc Fisher LSD test identified one at a time, no revisit was on average 2.5 to four minutes faster than the other versions. Interestingly, this study suggests that decreased flexibility results in faster completions times and no difference in scores with relation to a more flexible test that could be considered to simulate PPTs (Bodmann & Robinson, 2004). In summary, there have been evaluative studies of paper and computer mode effects on a wide range of tests. Studies demonstrate that the findings vary as much as the tests do. This could be related to the type of test, the format changes necessary to convert the test to a computerized format, or to the content of the test. While there have been studies to evaluate many other tests, to date there has not been a study to evaluate the differences and potential reliability and validity of curriculum based measures for math administered by the computer.

There are a number of potential variables that may account for the differences between computer and paper curriculum-based measures. Response time refers to the amount of time it takes to respond without any other intervening variables such as problem solving or generation of answers. This ability may be different for each individual and therefore interfere with the time it takes them to respond. This response time is also affected by the familiarity of making that type of response and motor skills. For example, a child may have very little exposure to computers and therefore their response time may be slower due to unfamiliarity with the keyboard. Conversely, a child may have difficulty with motor skills and as a result their response time with pencil and paper responding may be slower. Another possible cause for differences in scores could be automaticity of the task being performed. Some groups of students having differences in the instruction received or the amount of instruction and practice time received will differ in their ability to produce answers rapidly. When discussing alternate forms of tests especially alternate modes, it is vital to address presentation style. Typically paper and pencil forms of CBA/CBM math are presented in a row and column format with about 40-60 problems per page. There is also the ability to skip problems and go back and change answers. In the paper presentation the respondent has the ability to visually see multiple problems at a time(clusters of problems). Some advantage could be proposed if the respondent has the ability to solve the multiple problems at a time or begin solving the next problem in the cluster while writing the answer to the last problem. Depending on the computer format chosen one or more of these may be options or none of these abilities may be available. The presentation of the problems is also often different. The problems may be presented one at a time (each problem is presented individually and an answer must be given to move to the next problem), in scroll format (all problems are presented vertically and responder can scroll up and down the page, responses may or may not be changeable), or in columns and rows (identical to the paper format, but curser moves from one to the next with the possibility to go back and change answers or with no correction allowed). For the CBAs and CBMs, it is common practice to use a random generation program to develop probes regardless of the mode used. In these measures, the problems presented are not typically performance dependent. Consequently, the problems are chosen and presented randomly from a pool of items regardless of response style. There are a vast number of ways a probe could be presented and all of these factors need to be considered when comparing modes of testing and specifically evaluating whether computerized curriculum measures are psychometrically sound.

Computers as the solution

Reportedly many professionals use forms of curriculum based measures to evaluate progress, achievement level, or instructional knowledge. Therefore, it is beneficial and necessary to have an effective, quick, and accurate means of collecting this information. There are a number of problems with the paper and pencil form as already mentioned. A fast, reliable measure of math performance will assist in curriculum planning. For example, a teacher may determine that it is necessary for her to review the concept of subtraction with regrouping by looking at the scores of her class in that area. This measure could also be useful in identifying possible referrals or in determining problem areas for a particular student. For example, timed math probes may be given to a class with one covering each type of problem that should be learned up to this point. If a student or group of students show lower than average score in a particular area it may be possible for the teacher to provide them with a simple intervention to bring them up to speed with the rest of the class.

Using paper and pencil probes to do this has the potential to introduce a number of problems that could be reduced or eliminated by the use of a computerized measure of math performance. In order to use curriculum based measures an individual must develop and copy multiple versions of the tests, this can be time consuming and costly. By administering the math probes via the computer, the time consuming tasks of scoring, entering data for comparisons and the possibility of human error can be reduced. By using a computer to administer, it is also possible to have immediate results, graphs of the class, and intra-individual comparisons rapidly. This work is all performed by the computer and takes no teacher to complete. This can be accomplished using some of the computer programs available such as MBSP, keeping in mind that they first should be validated. The computer also establishes a record to be used to show progress, assist in referral identification, and to show completion of accountability standards. As previously stated CBM is used for a variety of functions so a discussion of the beneficiaries of this solution should be discussed.

Beneficiaries

The benefits of CBA/CBM procedures have also been demonstrated for the process of classwide screening and as an extension the method by which Responsiveness to intervention can be initiated. An approach currently used may be to screen all students of a school in the three basic areas, math basic facts fluency, oral reading fluency and comprehension, and written expression. This process could take approximately 30 to 45 minutes per class depending on the number of skills being measured. This procedure can establish a number of comparisons used to determine educational modification or individual intervention needs. In reference to the use of classwide screening for the RTI model, programs have used these methods as a preliminary step to begin the three tier model. This may be conceptualized as assessing all of the children in the school to establish school wide norms to use in determining the students meeting criteria for Tier 1. Examples of actions taken at tier 1 include classwide intervention or simple, quick individual or small group intervention conducted by teacher. The progress during either intervention would typically be monitored through CBA/CBM procedures focused to the particular skill that is lacking. As one can see the use of CBA/CBM procedures and its variations is vast therefore the evaluation and analysis of its merits is warranted. Further the methods mentioned inherit a number of disadvantages found to result with the use of CBA/CBM procedures (VanDerHeyden, Witt, & Gilbertson, 2007).

As eluded to with the classwide screening discussion, while there are vast benefits in the use of classwide screening for educational planning, response to intervention teams, and accountability, there are also disadvantages to the conventional use of CBA/CBM procedures. For all the advantages discuss thus far, the use of the computer to administer, maintain records and graph results is a logical solution. Classwide screening can be time consuming and require extra staff and specialized training of staff to administer. The same issues for scoring, production, and establishing results are all relevant with classwide screening as well. Universal school wide screening has recently become more widely used. An example of this is Screening to Enhance Educational Performance (STEEP) (VanDerHeyden, Witt, Naquin, 2003) which is an evidenced based model for improving services to all children and reducing the need for special education and other special services. Considering that all instruction is not equal, it is necessary to determine if a student is a normal learner but has not received adequate instruction. STEEP uses a process by which a child must pass through gates in order to be evaluated. These gates are designed to rule out children that are low-performing or low-achieving but do not actually have a disability. Giving a full evaluation as is typically done after a teacher referral is a lengthy process. STEEP through use of these gates or screenings, which take less than a hour, has reduced the referral rate for all students by 33% and for minorities by 50%. STEEP has been found to be more reliable than teacher referral. In 406 cases, teacher referral was correct only 19% of the time while the STEEP process was found to be correct 53% of the referrals (in saying a child has a problem when the child truly did have a problem; VanDerHeyden, Witt, & Gilbertson, 2007). Teachers were also less accurate in identifying children who did not have a problem. Special education placement is most common for the category of Learning Disablilty(LD), students with a LD constitutes over half of the special education population. The most frequently used model currently involves giving a student an intelligence test and an achievement test and determining if there is a severe discrepancy between where a child should be performing as expected by his intellectual functioning. Environmental factors should be ruled out but the tests given do not provide this information and it is too often not collected or incorporated (VanDerHeyden &Witt, 2005). By giving these tests it is not known that the child's teacher is providing inadequate instruction, that the child has missed a large number of days this year and therefore missed necessary instruction, or that the child had poor instruction for the last two years. These assessments are child-focused and negate the more likely explanation of the child's performance, lack of instruction (VanDerHeyden &Witt, 2005). A proposed alternative to this model is Responsiveness to Intervention (RTI), a formalized model of eligibility evaluation through STEEP and continued intervention and monitoring. The use of the computer for administration, scoring and interpretation would also save time in this type of a process. In 2004, President Bush signed into law the Individuals with Disabilities Education Improvement Act which differs from the previous law in one important aspect. Previously practitioners were encouraged to use the IQ-Achievement discrepancy model for Learning Disability (LD) eligibility determination. Now the law provides for practitioners to be able to use alternative methods to determine eligibility (Fuchs & Fuchs, 2006). Responsiveness to intervention(RTI) is the idea that students are identified as LD when their response to effective interventions is dramatically below that of their peers. The main assumption that supports the use of RTI is that it is able to differentiate between two explanations for low achievement: poor or lack of instruction and LD. Since this model focuses on the student's ability to perform at the same level as their peers, curriculum assessments are often used to determine the class or school level and to compare to the individual's performance in order to establish a criterion by which the individual can be evaluated for inadequate instruction or a LD (Fuchs, 2003). The use of computers to administer curriculum assessment as suggested could be beneficial to practitioners in their use of RTI. Computers could be used to administer interventions and assess progress as determined by the practitioner. For example, if a child is below their peers in the area of math and particularly in their ability to quickly answer previously instructed material. Then their fluency could be increased and checked using a computer program designed for curriculum-based assessments in the area of math. Computer programs could be adapted or developed for a wide range of abilities and interventions. By using computers in their administration, practitioners would experience the same benefits as mentioned above such as decreased scoring time, decreased development time, automatic graphing, and easier comparisons. Screenings as mentioned above are also used as an entry level procedure in some state and school districts. A large number of educational systems use some type of screening prior to entering school and then throughout. It is necessary for these screenings to follow certain guidelines to avoid some of the pitfalls often found when screenings are used. These screening programs should be used to determine if a student might need special services or further evaluation, determine levels of entering students to allow the teacher to plan the appropriate curriculum, and identify those that might benefit from early interventions. Both of the models mentioned above, STEEP and RTI meet the requirements for following these guidelines which include: chose appropriate goals, match assessment to goals, use psychometrically sound assessments, incorporate research findings, include input from caregivers and naturalistic observation, insure that appropriate follow-up assessment is ordered and that program evaluation occurs(Rafoth, 1997). RTI and STEEP are two programs that are widely used and are increasing in use. Curriculum-based assessments are appropriate to use according to the guidelines outlined by Rafoth and the use of computers to administer these widely given assessments would make the task more effective and efficient. Pre-referral and screening programs are essential to the education system and provide many benefits to the students served by these programs. These programs and many others would benefit from the use of computers to administer curriculum measures.

A study conducted comparing paper forms of unit tests in an introductory psychology class to a non-adaptive computer test designed to be equivalent. The computer test in its design addressed some of the common criticisms of computer adapted tests such as review questions, change answers, skip questions, and return to previously answered questions. Results showed that test scores were not significantly different between the computer-based test and the paper tests. This study also addresses the concerns of computer familiarity by administering computer attitude measures. The results of these measures indicated that the participants were comfortable and confident in their abilities on the computer. There are two important points to consider when interpreting these results in relation to the current study: 1) The test was to be completed (completely) in thirty minutes which is a longer time period than the current study utilizes. 2) This study did mimic the design ideas that the alternative computer format will in the current study following the APA guideline that computer tests should provide test takers the same opportunities that paper tests do (Mason, Patry, & Bernstein, 2001). The researchers of the current study have completely a series of studies in this area. These studies support the rational for the current study and show logical continuation of the literature in the area of computerized Curriculum based measurement. The initial study sought to evaluate computerized Curriculum Based Measurement using the procedures as set forth by Deno (1985) and Shinn (1989). Using a single basic math skills (addition), participants were administered multiple randomly generated paper and computer CBMs. The study also evaluated response time as a covariate. Response time refers to the ability to write or type numbers without the confound of calculation. Initial correlations of the multiple measures for both modes resulted in high internal consistency (paper & pencil = .951, computer = .894) suggesting that both types are reliable measures. Within the regression model, the multiple correlation coefficient was (.883) indicates a strong relationship between the two assessment types. The coefficient of determination, indicates that about 78% of the variation in computer performance is explained by the model and the addition of response time using the computer accounted for little of the variance. The regression model resulted in an equation for predicting paper performance given computer performance (paper = 1.413 x computer -8.107) indicating that although reliability and concurrent validity indexes were high, performance was consistently less when utilizing the computer. The next study in the series utilized the computer program to evaluate the progress effects of immediate feedback. Immediate feedback is only an option on the computer and can not be provided in the paper mode which makes this a distinct advantage of the computer mode. This study replicated the findings of the first study by evaluating the reliability of the paper and computer forms of CBM. It then through twenty practice sessions (total participants were divided randomly between the groups) evaluated the growth of each of three formats: paper, computer no feedback, and computer with feedback. All sessions were conducted according to the guidelines as set forth by Deno (1985). The skill used with subtraction from 18 because this skill was one that all of the children had been identified as needing improvement. Immediate feedback was administered on the top of the computer screen in a sentence format stating whether the last answer was correct or incorrect and then the correct response to the problem. For the problem 18 – 4, if the child answered incorrectly the computer response would read, “That is incorrect, 18 – 4 = 14. If the child answered correctly, the computer response would read, “That is correct, 18 – 4 = 14. Also if the answer was correct the text would be presented in green and if incorrect the text would be red. Results of this study were interpreted as pretest and posttest scores to show growth. All groups were administered the paper and computer pretest and posttest. The growth on the paper posttest was 28%, 9%, and 11% for paper, computer no feedback, and computer feedback, respectively. On the computer posttest growth was 34%, 39%, and 49%, respectively. This data would suggest that growth was greatest for the computer with feedback group when tested on the computer. When tested on the paper the growth was greatest for the paper group. There are some speculation as to why this is the case. There may be two components accounting for the growth seen, accuracy and fluency of automatic calculation and motor response. To say it another way, by practicing on the computer a student may increase accuracy and fluency of calculation and motor response of pushing numbers while practicing on the paper results in an increase in accuracy and fluency and motor response of writing numbers. Therefore without the practice of writing numbers when presented with the task of writing numbers as opposed to typing them the group that has practiced that skill will be superior. This study also found similar results for the reliability and validity of the computerized CBMs. This suggests that there is some factor affecting the computer which results in lower performance.

One factor that has been suggested in the literature is that of visual presentation and more restrictive response options for the computer versions of tests. When Mason et al. found no significant difference between the paper format and the computer format of multiple unit test for introductory psychology it was using a test that was designed to include the options that a paper tests has as equally as possible. These characteristics include seeing multiple items at once, changing answers, skipping problems, and returning to past items. The current study seeks to use this concept as the rational for its direction.

The purpose of the current study is to evaluate the reliability and validity of computerized CBM. The reliability of multiple versions of a basic skills probe for each mode will be a replication of the first study conducted by the authors. This will further support the reliability of computerized CBM by replicating the previous study with a different mathematics basic skill and a different age group. It is also of importance considering previous findings that an effort be made to modify the computer format in attempt to increase the equivalence of the two measures. This has been accomplished by providing the computer format with similar characteristic as the paper and pencil format. Considering these results, the concurrent validity of computerized CBM will again be evaluated with the addition of the evaluation of a new computer format and its concurrent validity with paper CBM. The new computer format includes more similar characteristics as the paper and pencil format.

The questions being asked through this study are as follows: 1) Is computerized CBM, Computer Paper or FlashCard, a reliable measure of performance on basic mathematics skills for elementary students? 2) Are there significant differences between the three modes of administration (paper, computer paper, flashcard)? 3) How does grade level affect the differences in performance between modes of administration?
It is hypothesized that both Computer Paper and FlashCard modes of computerized CBM will be reliable measures of basic mathematics skills. It is also hypothesized that the new Computer Paper mode of computerized CBM will correlated more highly with the paper

and pencil CBM. The Computer Paper format will reflect greater concurrent validity with paper and pencil format. Finally it is hypothesized that grade level will differentially affect the differences found between modes of administration.

CHAPTER III

METHODOLOGY

Participants

Participants were third, fourth and fifth grade students from general education classrooms in an elementary and intermediate school in a rural public school district. Students from 5 third, 4 fourth and 5 fifth grade classes were solicited for participation in the study by means of consent and assent forms for parents and students. Students and their parents who returned a signed informed consent were included in the study (See Appendix A, B for consent forms). All participants were treated in accordance with standards approved by the University Institutional Review Board for research with human subjects.

Materials

Computer based and traditional paper based versions of a multiplication probe were developed for use in this study. The computer based probes were designed to provide a mechanism for administering math probes via the computer to large groups of students at one time. The computer based probes also allowed for automated calculations of performance and input in a database for quick, efficient analysis. The paper probe consisted of multiplication fact problems with products to 81 presented in a vertical format. The multiplication problems consisted of randomly generated numbers from 0 to 9 presented as multiplication fact problems in a vertical format using an Excel® spreadsheet. Each student responded to the multiplication problem by writing the answer below the presented problem in a space provided. The spreadsheet was created in such a way to allow generation of multiple versions of the probes. There were two versions of the computer multiplication probe. Both computer based multiplication probes utilized a web based Javascript program that randomly generated multiplication problems with products up to 81. The generated problems were presented vertically and input from the keyboard (using either the number pad or the main keyboard numbers) was necessary to respond to the generated problem. In the FlashCard version of the computer probe, the problems were presented one at a time, with the next problem being presented only after a response is provided. This version would not allow for a responder to skip problems or return to problems once they had been responded to. This form was referred to as the FlashCard version. The other version of computer probe presented problems in rows and columns similar to that of the paper format. Problems could be navigated by mouse or arrows keys but the cursor would automatically move to the next problem after a response was given. This format provided respondents the ability to skip and go back to problems. This format is referred to as the Computer Simulated Paper version. A timer and a data collection mechanism that allowed for the transmission of student performance to a secure data base was built into both web based programs. The multiplication facts were basic skills that the participants had previously been or were currently being instructed upon.

Design of the Study

The current study was designed to examine the differences between method of assessment of curriculum based measurement administration. To accomplish this, the study was designed to examine performance as a within subject variable because all participants received all methods of assessment. The study also examined the differences that might be accounted for or altered by the change in grade status. Therefore grade was the between subjects variable and had three levels: third, fourth, and fifth grade.

The dependent measure for this study was digits correct per minute as a result of the two minute probes completed. A correct digit was determined by the number of digits being written or typed under the line in the correct place value. For example, for the multiplication problem 4 X 5 = the correct response was a 2 first and then a zero for the response of twenty. This response would result in a score of two digits correct for this problem. The number had to be in the correct order therefore 0 then 2 is incorrect. In this case a score one digit correct could be obtained by having either the two or the zero in the correct place value and another number in the other place holder.

Procedure

This study assessed students on multiplication performance as measured by paper and pencil and two versions of computerized probes. The student was administered three each of the multiplication probes to allow the researcher to establish the measure as reliable. Prior to implementation of the study, researchers met for a training session to familiarize themselves with the programs and the equipment. Researchers asked questions and were provided with feedback. During the study, students were randomly assigned to a seat in the computer lab. Researchers directed students to their seats. Instructions were given as a group to the students to clarify the procedures (Appendix C-script). Participants were asked for questions and researchers checked for understanding. Probes were administered in a convenient randomized order. Three paper and pencil timed single skill probes consisting of multiplication problems, three FlashCard computerized timed single skill probes consisting of multiplication problems (http://fp.okstate.edu/duhong), and three Computer Simulated Paper computerized timed single skill probes consisting of multiplication problems (http://fp.okstate.edu/duhong) were administered. Computation probes were administered in one of six sequences (ABC, ACB, BAC, BCA, CAB, CBA) with paper and pencil being represented by "A", Computer Simulated Paper based being represented by "B", and FlashCard based being represented by "C". The participant's sequence assignment was posted on the back of their chair to aid researchers in administration. All assessments were two minutes in duration. For assessments requiring paper, the administrators placed probes face down in front of the participant. Participants were asked to write their first and last name on the back of the paper. They were then instructed on how to complete the worksheet and asked for any questions (See script in appendix C). The administrator then instructed the participants to begin and they flipped the paper over and completed the problems provided. After two minutes the participants were told to put their pencil down and hold their paper in the air to be picked up. Assessments requiring computer administration were accessed via the web. With the assistance of the researcher the participants accessed the program to begin administration. At the appearance of the first problem or set of problems the students began responding to the problems. Participants were automatically stopped after two minutes when a THANK YOU screen which automatically appeared on the computer screen. No feedback about performance occurred at this time. After all probes were administered, students were returned to class. All participants were pulled out of their class as a group for 30 minutes to an hour. There was some variation in total time commitment for students due to the number of students that participated in their class. The time variations were due to increased time needed for researchers to direct students to the next method of assessment with a larger class.

CHAPTER IV

FINDINGS

This study was 3x3 mixed factorial design. Students who performed at less than an 80% accuracy level were eliminated from analysis to prevent the interference of accuracy on the results. Presenting the results before elimination provided insight into how inaccuracy affected fluency performance on the various modes.

It was hypothesized that both Computer Simulated Paper and FlashCard methods of assessment for computerized CBM would be reliable measures of basic mathematics skills. Correlations were calculated to evaluate test-retest reliability of the computer administered CBM probes for each method of assessment of computer. As a comparison, correlations were calculated for the Paper because it is considered to be the standard method of assessment. Correlations for the Paper averaged .87 and ranged from .82 to .93. Correlations for the Computer Simulated Paper averaged .92 and ranged from .89 to .94. Correlations for the FlashCard averaged .93 and ranged from .92 to .95.

It was also hypothesized that the new Computer Simulated Paper of computerized CBM would be correlated more highly with the paper and pencil CBM. It was suggested that the Computer Simulated Paper format would reflect greater concurrent validity with paper and pencil format than the previously tested Flashcard format. Finally it was

was hypothesized that grade level would differentially affect the differences found between formats of administration. A 3x3 split plot factorial (repeated measures) was employed to test for differences between the three formats of assessment for CBM. Grade level represented the between subjects variable and had three levels; third, fourth and fifth grade. The method of assessment represented the within subject variable and also had three levels; Paper, Computer Simulated Paper, and FlashCard. The dependent variable was fluency performance measured by digits correct on each probe. This statistical analysis was used because the design required repeated measures of a similar type. This analysis accounted for the multiple error terms and adjusted for the repeated measures. Due to the significant elimination of subjects that resulted from following the 80% accuracy rule, the sample was examined prior to elimination first. There was a significant interaction, $F(4,260) = 6.41$, $p < .001$, $\eta^2 = .09$, such that participants performance on the assessments varied based on grade level and assessment format. There was a main effect of grade, $F(2,130) = 75.53$, $p < .001$, $\eta^2 = .28$, such that fifth graders obtained more digits correct than fourth graders and fourth graders obtained more digits correct than third graders. There was also a significant main effect for format, $F(2,260) = 49.19$, $p <.001$, $\eta^2 = .54$, such that there were significant differences between performance on the various formats. Due to multiple level of each variable, it was necessary to examine simple effects and simple contrasts to determine the location of the significant differences. Significant differences were found for all grades between all methods of assessment with the exception of two comparisons in the third grade. A comparison of paper to computer simulated paper for third graders yielded a significance of $p = .077$. A comparison of computer simulated paper to flashcard yielded for third graders yielded a significance of $p = .753$. Table 1 displays all simple contrasts and Figure 1 shows the performance of all grades with regard to format.

In order to observe the effects of inaccuracy in this analysis those students who were less than 80% accurate were eliminated from analysis. Fifty students were eliminated due to inaccuracy. A significant interaction was demonstrated, $F(4,160) = 4.65$, $p = .001$, $\eta^2 = .10$, such that participants performance on the assessments varied based on grade level and assessment format. There was a main effect of grade, $F(2,80) = 33.17$, $p < .001$, $\eta^2 = .36$, such that fifth graders obtained more digits correct than fourth graders and fourth graders obtained more digits correct than third graders. There was also a significant main effect for format, $F(2,160) = 44.84$, $p <.001$, $\eta^2 = .45$, such that there were significant differences between performance on the various formats. Due to multiple levels of each factor, it is necessary to examine simple effects and simple contrasts to determine the location of the significant differences. Significant differences were found for all grades between all modes with the exception of two comparisons. A comparison of paper to computer paper for third graders yielded a significance of $p = .080$. A comparison of computer simulated paper to flashcard yielded for fourth graders yielded a significance of $p = .067$. Both of these almost reach significance indicating that the performance on the three methods of assessment is often significantly different regardless of grade. Table 2 displays all simple contrasts and Figure 2 shows the performance of all grades with regard to mode.

For third grade, means for performance are as follows: Paper(22.31), Flashcard(11.11), and Computer Simulated Paper (16.53). For fourth grade, means for performance are a follows: Paper(42.38), Flashcard(35.84), and Computer Simulated

Paper (32.84). For fifth grade, means for performance are as follows: Paper(73.85), Flashcard(63.65), and Computer Simulated Paper (57.72). These means reflect students that met criteria for accuracy. See Tables 1 and 2 for complete means for both analysis of all students and all accurate.

CHAPTER V

CONCLUSION

The study of different formats of curriculum based assessment that are currently being used is essential to provide support for their use. Children are referred for assessment and intervention with a variety of educational difficulties. The instruction provided by teachers is just as varied. It is helpful to be able to assess groups of children quickly and to have confidence that the results are concise. It should be noted that computerized curriculum based assessment is currently being used in practice but lacks uniformity and norms as well as empirical support for its use. As stated previously the benefits of using the computer to administer assessments in the case of curriculum based measurement include efficient administration, uniform administration, rapid results, and interpretive assistance.

The current study employed a 3x3 repeated measures design to examine the differences between formats in math curriculum based measurement. Participants were multiple classes of third, fourth and fifth graders at a rural elementary and intermediate school. One hundred and thirty three students participated with parental consent. Due to high rates of inaccuracy, analyses were completed for all children included in the study and then again on all students who performed above 80% accuracy. When the elimination of the inaccurate students was complete, fifty students were eliminated. This study was designed to answer three main questions: 1) Are computerized CBM, Computer

Simulated Paper and FlashCard, reliable measures of performance on basic mathematics skills for elementary students? 2) Are there significant differences between the three modes of administration (paper, computer simulated paper, flashcard)? 3) How does grade level affect the differences in performance between formats of administration? First the data suggested that the two computerized formats demonstrated strong concurrent validity with the paper format. Concurrent validity refers to when a measure is strongly correlated with a measure that has been previously validated. While this is the case, there are significant differences between performance on the three formats of administration as well as the three grade levels. The computer formats that were developed and used for this study were not found to be equivalent to paper and pencil method.

While the data were different when comparing the all subjects versus only the accurate participants, most of the comparisons did reach significance. Therefore further examination of the assessment methods is necessary to support the use of computerized assessment in its variety of capacities. Significant differences for performance were found for both assessment method and grade level. There was also a significant interaction between assessment method and grade level. As a result simple main effects and simple comparisons were performed to pinpoint the location of the differences within the levels of assessment method and grade level. When examining the pairwise comparisons, significance varied slightly for the two analyses. For the analysis including all children, all comparisons were significantly different with the exception of two: paper to computer paper and computer paper to flashcard for third graders. For the analysis including only accurate children all comparisons were significantly different with the exception of two: paper to computer simulated paper for third graders and computer simulated paper to flashcard for fourth graders. See Table 3 and Table 4 for pairwise comparisons. Regardless of accuracy levels of the participants, most of the comparisons reached significance. When examining the means, there is important information with regard to directionality. See Tables 1 and 2 for complete comparison of means. When looking at accurate students, third graders performed higher on paper measures than either computer measure. This may indicate that students that are in the early stages of learning a new skill will be less able to generalize their skills to an unfamiliar mode. For fourth graders there was significant difference between Flashcard and computer simulated paper performance. This difference may be due to an increased ability to generalize skills to computer performance for fourth grade with some limitation of performance on the computer simulated paper mode. These explanations are simply speculation and more research is needed to examine the reasons for these differences in performance across format and grade.

This study did indicate that computerized CBM is systematically related to paper CBM. All three formats demonstrated excellent reliability and could be used to assess progress. The computerized format is especially teacher friendly for use as a screening tool due to its quick administration in groups and efficient and accurate scoring and graphing of results.

While further research is necessary to determine the cause of differences found in this study, some possibilities may be speculated. Due to the slight increase in differences between format as the education level increased, it could be suggested that computer familiarity played a factor. Since there is not a standardized measure of computer familiarity for children, this construct may be difficult to assess. In previous research, response time was examined in attempt to test this theory. In that study response time accounted for only 2% of the variance between a paper format and computer flashcard format (Duhon, Wong, and Mesmer, 2007). It can also be extrapolated that fluency may influence differences. This could be examined at multiple ways, as children become more fluent the differences in mode increase, as fluency increases on the paper (practice method of choice) differences between paper and computer modes increase, or there is a difference between the nonmathematical abilities required to complete the task fluently. There may also be stimuli that are present when completing a computerized curriculum based measure that are not present during paper administration. Assuming that the majority of children have been exposed to paper administration more frequently than computer administration, there is reason to hypothesize that children have increased familiarity with paper administration stimuli. These possibilities should be studied in additional research.

Even though, the computerized measures were not equivalent to the paper mode there are important implications. Each assessment method was reliable and therefore any of the modes could be used as a method to monitor fluency and measure growth. Further research should be conducted to evaluate the computerized forms for sensitivity to growth. Using computerized administration instead of paper is beneficial due to conservation of resources. The resources conserved would include paper, teacher time for administration and teacher time in scoring and evaluation. Additionally, the computerized modes after examination of rank ordering could be used for screening individuals who need fluency intervention and progress monitoring. Rank order was maintained in the present study, i.e. regardless of mode the students performed in approximately the same rank order for each class. This can be an important tool to use in identifying students who are in need of additional intervention for certain skills. For example, a school may decide to implement an additional program to insure that all students can perform at determined level prior to state testing. Using computerized CBM, a school can efficiently screen the entire school and select the lowest performing five percent in each class to for intervention. Following the intervention the school can again use screening to evaluation effectiveness of the intervention and insure that the classes meet requirements. An individual teacher may follow the same concept to identify children of more individualized programs for any number of skills within her curriculum. To make the current results more helpful, a future step in this area of research would be to attempt to identify and eliminate the differences between the paper and computer mode. Differences may include the physical differences that are required to perform using each mode. In elementary school one of the first skills taught is how to hold a pencil and then how to write using that pencil, crayon or marker. While advances in technology have promoted increasingly earlier introduction to computers, the method of choice for academic work remains the pencil. Therefore the pencil mode is arguably most familiar. The question now is how much practice it will take to eliminate the physical practice effects that make paper different than computer. This concept that the physical response of paper is faster than the physical response of computer is puzzling to some since the response of pressing a key seems to be faster than writing a number or two. At closer examination, when a child is not familiar with a keyboard finding a number or two on the keyboard may take significantly more time. To examine another physical difference, the process necessary to enter an answer on the computer may also extend the response time. There is a visual change that has to take place: the child must look at the computer to see the problem then must look at the keyboard to find the number to answer the problem. Whereas in the paper format, the child is able to continuously look at the paper to identify the problem and respond. There are also likely a multitude of other differences that will need to be examined and if possible eliminated between the two modes.

A limitation of this study is that the researchers were unable to collect data of computer experience which may have been helpful in understanding differences in performance. There were also some computer malfunctions which resulted in six children being eliminated from the study due to incomplete data sets. Limitations that are specific to the computer paper mode are that the students were required to enter the numbers in a certain order that they may not have been used to resulting in errors that may have not been true errors. Additionally for computer paper the students had to tab or used the mouse to move from one problem to another if they did not enter the correct answer to the previous problem. This may have caused delays. Finally the skill chosen for this study may have been too advanced for the subjects used and using a skill that was known to be introduced to all students may have yielded different results. This last limitation is unlikely due to the differences found for the fourth and fifth grades of this study but it is a variable to consider in future studies.

Future studies may include changes to the current computer formats. A change that may be helpful in eliminating differences in the flashcard format includes making the program so that children are able to skip when they don’t know a problem rather than having to enter what they know is an incorrect answer to move on. Changes for the computer paper format may include larger problems or practice with the computer paper format to increase familiarity with format. Future work needed in this area of research is the development of norms for a computerized format. Developing computerized norms would eliminate the need to demonstrate equivalence of computer and paper performance as well as support the use of computer assessment for CBM which is already in practice. The effect of practice using computer formats is another area that would benefit from future research. A study may look at the amount of practice necessary to eliminate the differences between a computerized format and the paper format. This would be helpful in determining if practice with a computerized format for two weeks would than enable a teacher to use computers the measure progress throughout the year. In the comparisons of other computer assessment with their own paper format, some studies have also compared computer skill assessment to the performance demonstrated on a standardized assessment. In some cases these studies have been able to suggest a predictability of performance on standardized measures based on computer performance. These two research ideas should be considered for future studies. Finally as stated previously the use of rank ordering may be beneficial in identifying students at-risk and in need of intervention. A study demonstrating these benefits would greatly contribute to the computerized CBM literature.

REFERENCES

Allen, M. J. & Yen, W. M. (1979). Introduction to Measurement Theory. Prospect Heights, Illinois: Waveland Press, Inc.

Allinder, R. M., Fuchs, L. S., & Fuchs, D. (1998). Computer-Assisted Assessment. In H. B. Vance (Ed.), *Psychological assessment of children* (pp.87-105). New York: NY: John Wiley & Sons, Inc.

Allinder, R. M., Fuchs, L. S., & Fuchs, D. (1998b). Curriculum-based measurement. In H. B. Vance (Ed.), *Psychological assessment of children* (pp.106-129). New York: NY: John Wiley & Sons, Inc.

Bodmann, S.M. & Robinson, D.H. (2004). Speed and performance differences among computer-based and paper-pencil test. *Journal of Educational Computing Research, 31*(1), 51 – 60.

Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educastional Technology, 33,* 593-602.

Cundari, L.A. & Suppa, R.J. (1988). The potential uses of curriculum-based assessment for decision-making in special education. *Exceptional Child, 35(3),* 143-154.

DeAngelis, S. (2000). Equivalency of computer-based and paper-and-pencil testing. *Journal of Allied Health, 29,* 161-164.

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

Deno, S.L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36,* 5-10.

Deno, S. L. (2003). Curriculum-based Measures: Development and perspectives. *Assessment for Effective Intervention, 28*, 3-12.

Deno, S.L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Duhon, G., Wong, C., & Mesmer, E. (2007). Evaluation reliability and validity of a computerized version of CBM. Poster presented at the National Association for School Psychology conference, New York, NY.

Fuchs, L. (2003). Assessing intervention responsiveness: conceptual and technical issues. *Learning Disabilities Research and Practice, 18(3),* 172-186.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique, 8,* 135-149.

Fuchs, D. & Fuchs, L.S. (2006). Introduction to Response to intervention: What, why and how valid is it? *Reading Research Quarterly, 41(1),* 93-99.

Fuchs, L. S., Fuchs, D. & Courey, S.J. (2005). Curriculum-based measurement of mathematics competence: From computation to concepts and applications to real-life problem solving. *Assessment for Effective Intervention, 30,* 33-46.

Fuchs, L. S., Fuchs, D., & Hamlett, C.L. (1989). Effects of instrumental use of Curriculum-Based Measurement to enhance instructional programs. *Remedial and Special Education, 10(2),* 43-52.

Fuchs, L. S., Fuchs, D., & Hamlet, C.L. (1994).Strengthening the connection between assessment and instructional planning with expert systems. *Exceptional Children, 61*, 138-146.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1989). The reliability and validity of skills analysis within curriculum-based measurement. *Diagnostique, 14*, 203 – 221.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review, 20,* 49-66.

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*, 617-624.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Monitoring basic skills progress: Basic spelling* [computer program]. Austin, TX: Pro-Ed.

Fuchs, L. S., Hamlett, C. L., Fuchs, D., Stecker, P. M., & Ferguson, C. (1988). Conducting curriculum based measurement with computerized data collection: Effects on efficacy and teacher satisfaction. *Journal of Special Education Technology, 9,* 73-86.

Gay, L.R. & Airasian, P. (2003) Educational Research: Competencies for analysis and applications (7th Ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *Journal of Special Education, 36*(2), 102-112.

Howell, K. W., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education.* Columbus, OH: Charles Merrill.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1996). *Models of curriculum-based assessment: A blueprint for learning* (2nd ed.). Austin, TX: Pro-Ed.

Kazdin, A.E. (2003). *Research Design in Clinical Psychology*(4th ed.). Boston, MA: Allyn & Bacon.

Layton, C.A. & Lock, R.H. (2007). Use authentic assessment techniques to fulfill the promise of no child left behind. *Intervention in School and Clinic*, *42*(3), 169-173.

Marston, D. B. (1989). A Curriculum-Based Measurement Approach to Assessing Academic performance: What is it and Why Do It. In Shinn, M. R. (Eds.), Curriculum-Based Measurement: Assessing Special Children (pp. 18-78). New York, NY: The Guilford Press.

Mason, B. J., Patry, M., & Bernstein, D.J. (2001). An examination of the equivalence between non-adaptive computer-based and traditional testing. *Journal of Educational Computing Research, 24(1),* 29 – 39.

Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114,* 449-458.

Rafoth, M.A. (1997). Guidelines for developing screening programs. *Psychology in the Schools, 34(2),* 129-142.

Shapiro, E.S., Angello, L.M., &Eckert, T.L. (2004). Has curriculum-based assessment become a staple of school psychology practice? An update and extension of knowledge, use, and attitudes from 1990-2000. *School Psychology Review, 33,* 249-257.

Shapiro, E.S., Keller, M.A., Lutz, G., Santoro, L.E., & Hintze, J.M. (2006). Curriculum-based measure and performance on state assessment and standardized tests: reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*, 19-35.

Shinn, M. R. (Ed.) (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

Shinn, M. R. (Ed.) (1998). *Advanced applications of curriculum-based measurement.* New York: Guilford Press.

Shinn, M. R. & Marston, D. (1985). Differentiating mildly handicapped, low-achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*, 31-45.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is Measured in Mathematics Tests? Construct Validity of Curriculum-Based Mathematics Measures. *School Psychology Review, 31*, 498-513.

Tindal, G. (1992). Evaluating instructional programs using curriculum-based measurement. *Preventing School Failure, 36*, 39-42.

Tindal, G. & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities: Research & Practice, 6*, 211-218.

Tindal, G., Marston, D., & Deno, S. (1983). The reliability of direct and repeated measurement (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

Turnbull III, R. H. (2005). Individuals with disabilities education act reauthorization: accountability and personal responsibility. *Remedial and Special Education*, *26,* 320-326.

VanDerHeyden, A.M. & Witt, J.C. (2005). Quantifying context in assessment: capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34(2)*, 161-183.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Reponse to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225-256.

VanDerHeyden, A.M., Witt, J.C., & Naquin, G. (2003). Development and Validation of a process for screening referrals to special education. School Psychology Review, 32(2), 204-227.

Witt, J. C., Elliot, S. N., Daly III, E. J., Gresham, F. M., & Kramer, J. J. (1998). Assessment of at-risk and special needs children. (2nd ed.). Boston, MA: McGraw.

APPENDICES

Appendix A: Consent Letter

Parent Permission Form

Research Project title: Three modes of Delivery for Curriculum Based Measurement Math

Dear Parent,

This form is to request permission for your child ___________________________ to participate in a research study conducted by Cassie Wong, a doctoral student from Oklahoma State University. The research project will take place at ___________________ Elementary during regular school hours. Your child will be taken out of class for approximately 30 to 45 minutes on one day the____________________. They will be taken to the computer lab where they will be asked to complete several math worksheets similar to the ones they have completed in regular class. The worksheets will consist of all simple multiplication problems. They will also be asked to complete the same type of math problems on the computer. The purpose of this study is to evaluate the equivalence of paper and computer administration of curriculum based measures for math. The study includes two different formats of computer administration to determine the most equivalent format. Your child or yourself can withdraw your child's participation at any time during the study without any penalty.

Your child's teacher will be informed of the results of this study to assist them in their instructional planning with your permission. If your child's results are included in any research reports, his or her name will not be included in the report. In fact, no information that would result in your child being personally identified such as the school he/she attends or the state or town where he/she lives will be revealed. Every effort will be made to maintain the confidentiality of the data obtained from this study. The data will be housed at Oklahoma State University and only the Principal Investigator and the research assistant working on the project will have access to it. Data collected regarding your child will be stored in a locked file cabinet in the Principal Investigators office. Further, these data will be entered into an electronic database on a computer that is password-protected program. Specific access to data within the computer will also be password protected and will only be available to the Principal Investigator and the research assistant working on this project.

If you have any questions please contact Cassie Wong (405) 614-1428 or Dr. Gary Duhon (405) 744-9436. If you have questions about your rights as a research volunteer, you may contact Dr. Sue Jacobs, IRB Chair, 219 Cordell North, Stillwater, Ok 74078, 405-744-1676 or irb@okstate.edu.

________Yes, I give my permission for my child to participate in this study and for researchers to share the results of this study with my child's teacher for instructional planning purposes.

________Yes, I give permission for my child to participate in this study but do not give permission to the researchers to share the results of the study with my child's teacher.

________ No, I prefer that my child does not participate in this study.

______________________________ ______________

Parent signature Date

Appendix B: Assent Letter

Child Assent Form

Research Project title: Three modes of Delivery for Curriculum Based Measurement Math.

Principal Investigator: Cassie Wong
Phone Number of Investigator: (405)614-1428

Read the following sections to the student.
Purpose:
I want to see how well you can work multiplication problems.

Procedures:
Children who participate in this project will be working math problems on the computer and on paper. You will be working multiplication problems. This project will include you completing several worksheets. You will be working some problems on the computer and some on paper. You have two minutes each time to working as many problems correctly as you can. Please try to do your best work.

Risks:
My working with you will not change what you and your teacher already have planned, so all that will change is that I am going to be asking you to work some math problems.

Alternative Procedures.
You do not have to be involved in the study if you do not want to. You can stop at any time you want. You do not have to do anything that makes you feel uncomfortable or sad. No one will be upset with you if you say “no” or if you say “yes” and then change your mind. You can change your mind at any time.

You have been told about the study.
You have been told that you do not have to do any of the tests if you do not want to.
You have also been told that you can stop any time you want, even after you begin.

I agree to participate

____________________________________ ____________________
Signature of Child Date

____________________________________ ____________________
Signature of Person Reading and Obtaining Consent Date

Appendix C: Script

Instructions for computer and paper math probes simultaneously

1. Direct students to previously randomly assigned seats as they enter the room. Orders of administration will be posted to the back of their seat.

2. Pass out and allow student to review and sign assent. Retrieve assent forms and check for signatures. Inform students that if their order starts with an A they will be doing paper first and if they have a B or C first they will start on the computer.

3. Say, " 'A' students, I am passing out a worksheet, it will consist of mixed mathematics problems. Please leave it face down and write your number on it. 'B or C' students sign in on the computer under your number. You should now see a large orange square. 'A' students when I say start you will flip your paper over and begin working the problems. Work as quickly as you can without making mistakes. 'B or C' students, when I say start you will click the orange square and begin working problems as quickly as you can without making mistakes.(demonstrate computer format for all students before beginning) The computer screen will stop presenting problems after two minutes. Again work as quickly as you can without making mistakes. 'A' students, after two minutes I will instruct you to stop, put you pencil down, and hold your paper in the air. You will be completing multiple worksheets. Are there any questions?

4. Ready? Start.

5. After two minutes, Say "Paper group Stop, put your pencil down, and hold your paper in the air." Wait until all students have stopped and instruct them to change modes. Repeat until each group has completed all probes.

6. Debrief and dismiss students.

Table 1.

Table of Means and Standard Deviations (All Students)

Grade	PP	SD	SP	SD	FC	SD
3	13.30	(8.97)	10.11	(6.89)	9.73	(6.22)
4	36.40	(18.43)	28.94	(15.33)	31.56	(16.94)
5	67.43	(29.69)	53.73	(22.69)	60.11	(24.42)

Note. PP-paper and pencil, SP-simulated paper, FC-flashcard

Table 2.

Table of Means and Standard Deviations(All accurate students)

Grade	PP	SD	SP	SD	FC	SD
3	22.31	(5.73)	16.53	(5.37)	11.11	(5.60)
4	42.38	(19.54)	32.84	(16.75)	35.27	(19.26)
5	73.85	(28.10)	57.72	(22.41)	63.65	(24.94)

Note. PP-paper and pencil, SP-simulated paper, FC-flashcard

Table 3.

Simple Contrasts (All Students)

Grade	Format	Format 2	Mean Difference	Significance
3	PP	SP	3.181	.077
3	PP	FC	3.562*	.040
3	SP	FC	.381	.753
4	PP	SP	7.461*	.000
4	PP	FC	4.842*	.001
4	SP	FC	-2.618*	.008
5	PP	SP	13.698*	.000
5	PP	FC	7.318*	.000
5	SP	FC	-6.380*	.000

Note. *p<.05, PP-paper and pencil, SP-simulated paper, FC-flashcard

Table 4.

Simple Contrasts (All Accurate Students)

Grade	Format	Format 2	Mean Difference	Significance
3	PP	SP	5.778	.080
3	PP	FC	11.194*	.000
3	SP	FC	5.417*	.018
4	PP	SP	9.543*	.000
4	PP	FC	7.114*	.000
4	SP	FC	-2.429	.067
5	PP	SP	16.130*	.000
5	PP	FC	10.204*	.000
5	SP	FC	-5.926*	.000

Note. *p<.05, PP-paper and pencil, SP-simulated paper, FC-flashcard

Oklahoma State University Institutional Review Board

Date: Friday, August 10, 2007

IRB Application No ED0767

Proposal Title: Three Modes of Delivery for Curriculum Based Measurement Math

Reviewed and Processed as: Expedited (Spec Pop)

Status Recommended by Reviewer(s): Approved Protocol Expires: 8/9/2008

Principal Investigator(s

Cassandra Wong 3705 N. Monroe St. Stillwater, OK 74075	Gary J Duhon 423 Willard Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☑ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Sue C. Jacobs

Sue C. Jacobs, Chair
Institutional Review Board

VITA

Cassandra Brenner Wong

Candidate for the Degree of

Doctor of Philosophy

Dissertation: RELIABILITY, VALIDITY, AND EQUIVALENCY OF A COMPUTERIZED CURRICULUM BASED MEASURE FOR BASIC MATH SKILLS

Major Field: Educational Psychology, Specialization in School Psychology

Biographical:

Personal Data: Born in Germany in March of 1982 to Charles and Dianna Brenner. Raised in Lawton, Oklahoma and various military bases in Germany. Raised with three brothers and three sisters.

Education:
Completed the requirements for the Master of Science in Educational Psychology at Oklahoma State University, Stillwater, Oklahoma in July, 2006.

Completed the requirements for the Bachelor of Science in Psychology at Cameron University, Lawton, Oklahoma in May, 2004.

Experience:
Completed internship at University of Tennessee Health Science Center Psychology Consortium, Memphis, Tennessee in August, 2009.

Completed Clinic based Practicum at School Psychology Center, Stillwater, Oklahoma in June 2008.

Completed School Based Practicum at Westwood Elementary, Stillwater, Oklahoma in May 2007.

Name: Cassandra Brenner Wong Date of Degree: August, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: RELIABILITY, VALIDITY, AND EQUIVALENCY OF A COMPUTERIZED CURRICULUM BASED MEASURE FOR BASIC MATH SKILLS

Pages in Study: 66 Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Scope and Method of Study: This study examined the performance of third, fourth, and fifth graders on curriculum based measurements for math. The study sought to compare performance on two different computer formats to the traditional paper and pencil format. There were 133 initial participants and 80 participants after elimination for inaccuracy. Participants completed three each of the Paper/Pencil, Simulated Paper, and Flashcard probes. All probes were administered in a group format in a computer lab.

Findings and Conclusions: A significant interaction was demonstrated, $F(4,160) = 4.65$, $p = .001$, $\eta^2 = .10$, such that participants performance on the assessments varied based on grade level and assessment format. There was a main effect of grade, $F(2,80) = 33.17$, $p < .001$, $\eta^2 = .36$, such that fifth graders obtained more digits correct than fourth graders and fourth graders obtained more digits correct than third graders. There was also a significant main effect for format, $F(2,160) = 44.84$, $p <.001$, $\eta^2 = .45$, such that there were significant differences between performance on the various formats. Due to multiple levels of each factor, it is necessary to examine simple effects and simple contrasts to determine the location of the significant differences. Significant differences were found for all grades between all modes with the exception of two comparisons. A comparison of paper to computer paper for third graders yielded a significance of $p = .080$. A comparison of computer simulated paper to flashcard yielded for fourth graders yielded a significance of $p = .067$. Both of these almost reach significance indicating that the performance on the three methods of assessment is often significantly different regardless of grade. All three formats demonstrated excellent reliability and could be used to assess progress. However, the computer formats that were developed and used for this study were not found to be equivalent to paper and pencil method.

ADVISER'S APPROVAL: Terry Stinnett, PhD