Louisiana State University
LSU Digital Commons

LSU Doctoral Dissertations | Graduate School

2005

Constructing a math applications, curriculum-based assessment: an analysis of the relationship between apllications [sic] problems, computation problems and criterion-referenced assessments

James E. Connell, Jr.
Louisiana State University and Agricultural and Mechanical College, jconnell@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Psychology Commons

Recommended Citation
Connell, Jr., James E., "Constructing a math applications, curriculum-based assessment: an analysis of the relationship between apllications [sic] problems, computation problems and criterion-referenced assessments" (2005). *LSU Doctoral Dissertations*. 2944.
https://digitalcommons.lsu.edu/gradschool_dissertations/2944

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contactgradetd@lsu.edu.

CONSTRUCTING A MATH APPLICATIONS, CURRICULUM-BASED ASSESSMENT: AN ANALYSIS OF THE RELATIONSHIP BETWEEN APLLICATIONS PROBLEMS, COMPUTATION PROBLEMS AND CRITERION-REFERENCED ASSESSMENTS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Mechanical and Agricultural
College in partial fulfillment
of the requirements
for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
James E. Connell Jr.
B.A., Temple University, 1997
M.A. Louisiana State University, 2001
August 2005

©Copyright 2005

James E. Connell Jr.

All rights reserved

Dedication

My father and mother watched as I received my bachelor's degree in Temple University's, Mitten Hall. It brought tears to my father's eyes. He died while I was attending graduate school at LSU. He was proud of my choice to go to college at 27, and even prouder that I did so well. More than that, he was proud of me regardless, and that has made all the difference. Completing this doctorate is as much for him as it is for me.

Acknowledgements

To Ralph, who admitted me into the honors program at the Community College of Philadelphia and helped me “find my way”, and Dennis who doubted my resolve and capacity, and Madeline who exposed me to great the art and inspired me to travel to Italy five years later to see “The School of Athens”, I will always be grateful, and Evan who inspired me to read great literary works and patronize the theatre, and Henry for his enthusiastic interest in history, and Martin for setting a lofty model.

To Joe, who gave me my first job working in the field of behavior analysis and Bud, who pushed me to question my political and social convictions and insisted that I look beyond the obvious.

To Phil, a great mentor, great “thinker” and great friend. It was Phil that lead me to behavior analysis and provided me with the opportunity to pursue my “dreams” by exposing me to the intellectual content I needed to explain my world.

To Susan, thank you for your assistance with data collection, assessment validation and welcoming me into her district.

To Joe, who provided me with the opportunity to mature as a person and a researcher while providing the right amount of guidance and motivation through this process. I would not be here without your assistance.

Finally, to my best friend and partner Chrissy. We met on this journey and have soared and struggled through many events along the way. You are my best companion and perfect compliment and have brought me the greatest joy in the world, our son. Love, loyalty and friendship always.

Table of Contents

Copyright ii

Dedication iii

Acknowledgements iv

Abstract vi

Introduction 1

Method 16

Results 31

Discussion 47

References 58

Appendix A. General Curriculum Standards Reported by Seven Randomly Chosen States and Their Definitions. 62

Appendix B. Math Applications and Computation Assessments. 63

Appendix C. Basic Skill Worksheet, Sums to 18. 123

Appendix D. Teacher Rating of Student Performance and Expected Final Grade. 124

Appendix E. Assessment Rating Profile. 125

Appendix F. Directions and Scripted Instructions for Single Skill and Multiple Skill Math and Maze Assessments. 126

Appendix G. Procedural Integrity Check List. 127

Vita 128

Abstract

Curriculum-based measurement (CBM) is a well established tool for formative assessment. Although CBM in mathematics could potentially be used for other purposes, such as prediction of state test scores, validity coefficients between CBM and state tests in mathematics have been moderate at best (Skiba, Magnusson, Martson, and Erickson, 1986; Martson, 1989; Putnam, 1989). The purpose of the present investigation was to develop and evaluate a set of math assessments designed to measure the type of application and problem-solving objectives required on state tests. The "application" type assessments constructed for this study combined characteristics of CBM, accuracy-based curriculum-based assessment (CBA) and criterion-referenced assessment (CRA). The new assessments were derived from state standards and matched to local district curriculum. The methodology involved obtaining validity coefficients for (a) traditional CBM assessments versus (b) newly developed assessments which incorporated applications/problem-solving with regard to "state tests" which served as a standard against which (a) and (b) above could be evaluated.

The assessments examined included (a) one single skill computation assessment, (b) one multiple skill computation assessment, (c) one maze reading assessment, (d) a newly constructed applications assessment, (e) Woodcock-Johnson III math subtests, (f) Louisiana Education Program for the 21st Century (LEAP), and (g) the Iowa Test of Basic Skills (ITBS). Additionally, the various assessments were investigated relative to teacher-based indicators including (a) students' final grades, (b) teacher report of year end performance, and (c) teacher preference of math assessments. Participants included 172 first to fifth grade regular education students who were administered the CBM/CBA/CRA assessments one month following the state tests. State testing included use of the LEAP in fourth grade and the ITBS in second, third and fifth grades. Results indicated that the newly developed application assessment exhibited a stronger relationship with the criterion assessments, with students' final grades, and teacher report of year-end performance. In addition, the application type assessments were preferred over the computation assessments by all teachers. Results and limitations are discussed with regard to the construction and use of an application-oriented CBM/CBA/CRA for users needing assessment which might combine the power of formative evaluation with the ability to accurately predict performance on state tests.

Introduction

Prelude

Curriculum-based assessment is a set of procedures using "direct observation and recording of student performance in the local curriculum as a basis for gathering information to make instructional decisions" (Deno, 1987). Of the four basic skill areas (i.e., reading, writing, spelling and mathematics), math may be the least represented in the literature and has only modest technical adequacy (Thurber, Shinn and Smolkowski, 2002). Though some math curriculum-based measurement investigations have included word problems in assessments developed for middle school students (Espin and Tindal, 1998), assessments for students with mathematics disabilities (Fuchs, Fuchs, Hamlett and Appleton, 2002), and to identify growth indicators for low-achieving middle school students (Foegen and Deno, 2001), the majority of math curriculum-based assessment procedures has focused on math computation (sometimes called operations) and number sense problems (see Shinn 1989; 1998 for reviews; Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek and Stecker, 1994) rather than applications problems (concepts, numeration, problem solving, measurement, geometry, data analysis and number sentences).

The lack of application problems as part of math curriculum based assessment may be a function of the differences between the accuracy model of curriculum-based assessment (Gickling and Havertape, 1981; Gickling and Thompson, 1985; Hargis, 1987), the criterion-referenced model (Blankenship, 1985; Idol-Maestas, 1983), and the fluency model (Deno and Mirkin, 1977). As such, a summary of the models follows, identifying the characteristics of each model and highlighting the differences.

Accuracy Model. First, the accuracy model (for simplicity this model will be referred to hereafter as CBA) evaluates outcomes in terms of percent correct. That is, the data is often reported as a percentage with the number correct divided by the total number of problems. The model has been described as an instructional model in that the assessment results lead to intervention recommendations. Frisby (1987) claims that the model is "task analytic" in nature in that assessment results are an analysis of task demands reflecting the instructional materials (i.e., local curricula) students are expected to learn. Third, the model is instructionally based and presumes that effective instruction will then be implemented given the assessment results. As such, teachers are expected to develop their own instructional materials to teach those areas where each student performed poorly. Hagris (1987) argues that as teachers develop good instructional material, students will achieve higher levels of performance. The accuracy model historically has assessed student performance indirectly during academic learning time (ALT). Gickling and Havertape (1981) suggested that the benefits of assessing during ALT are: 1) measurement of performance relevant to instruction and 2) the sensitivity of the assessment to improvement in student performance.

There is very little evidence supporting the psychometric adequacy of CBA, including content validity. The lack of technical adequacy has been attributed to the lack of a standardized procedure in the development of the assessment materials. Gickling and Havertape (1981) suggested that they were not concerned how data were collected and recorded, only that it be done. In doing so, teachers could then use the data, 1) as a screening tool, 2) for instructional planning, 3) to identify low performing students in specific content areas and 4) to monitor progress of those areas once an intervention was developed.

Screening and special education eligibility determination would occur once the student-specific and task-specific interventions had failed.

Criterion-Referenced Model. In the criterion referenced model (referred to as CRA from here on), the goal is to obtain, "… direct and frequent measures of a student's performance on series of sequentially arranged objectives from the curriculum used in the classroom" (Blankenship and Lilly, 1981). The goal of this approach is to link assessment to instruction and the local curriculum content standards. Developed to incorporate content beyond basic skills, the approach was designed to monitor student progress in short term and/or long term objectives. In other words, the assessments developed could provide formative and summative data. According to Blankenship (1985), development of a criterion-referenced assessment requires the sequential selection of specific items from the curricula that are then combined to make an assessment. Students are then tested for mastery in the content areas selected (e.g., in mathematics the assessment would include applications and computation problems). Thus, if mathematics were selected as a content area that educators were interested in monitoring using this method, then that process might proceed as follows. First, the district content standards would be considered such that items on the assessment would represent the standards (e.g., number sense, algebra, geometry, data analysis). Next, items would be selected sequentially from the curricula from each of those standards. One possibility for this process would be to choose items from the chapter tests, beginning with chapter one, and proceeding through to the last chapter.

The number of items from each standard should be a consideration. In order for the assessment to be comprehensive, it would need to contain items from the beginning, middle and end chapters. However, in order for the assessment to be easy to construct and administer, the assessment should be limited to as few items as possible.

To date, there are very few studies that have investigated this model and none were found investigating math performance. Hence, there are no technical data to report from the literature.

Fluency Model. Characteristics of the fluency based model called curriculum-based measurement (CBM) includes, 1) direct measurement of student performance within curricula, 2) brief time-based assessments (i.e., one to five minutes per assessment), 3) repeated administration using alternate forms yielding data that can be graphed to show dependent variable changes due to curricular or instructional adaptations (Frisby, 1987). Curriculum-based measurement differs from CBA and CRA in that one purpose of CBM is a direct assessment of student performance for special education eligibility determination (Shinn, 1989). Additionally, CBM has been used as a progress monitoring tool to monitor student performance across the year in early literacy skills (Good and Kaminski, 1996); design instruction (Shapiro, 1996); assess academic readiness skills in kindergarten students (VanderHeyden, Witt, and Naquin, 2001) and to document academic gains in basic skills (Martson, Deno, and Kim, 1995). Additionally, there is considerably more research supporting the use of CBM in terms of reliability of alternate forms, validity and sensitivity to instruction, though a majority of the literature investigates the technical adequacy of oral reading fluency rather than other forms of CBM such as math, spelling or writing assessment.

During CBM inception, considerable research and development was conducted to identify reliable, valid basic skill measures (Shinn, 1989). Stanley Deno and Phyllis Mirkin established a significant portion of this work in the late 1970s and early 1980s (Marston, 1989). Five steps were taken to accomplish this task. First, the extant literature was reviewed to identify possible measures. Second, the selected measures were reviewed with respect to previously established criteria. Third, criterion validity was established with those measures. Fourth, reliability coefficients were obtained. Fifth, the logistical elements of the measures were determined (i.e., assessment interval length and measurement domain size).

The process resulted in CBM assessments in a variety of core content areas. In reading, assessors counted the number of words correctly read from a basal reader or class word list per one minute of time. In written expression, assessors counted the number of words a student wrote per three minutes. In math, assessors counted the number of digits correctly answered on grade-level computation problems per two minutes. Later investigations have increased math CBM assessment time to five minutes (Fuchs, Fuchs, Hamlet and Stecker, 1991) and more recently to eight minutes (Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek and Stecker, 1994). Finally, in spelling, assessors counted the number of words correctly spelled or correct letter sequences per two minutes. Spelling words were dictated every seven seconds (Shinn, 1989).

The above list of measures does not necessarily describe the inclusion of local curricula. In other words, it is not necessary to review the content standards, or curricula to develop reading, writing, spelling and math assessments that monitor student performance on basic skills. In fact, studies in reading have demonstrated that grade-level oral-reading assessments could be replaced by grade-level, cloze and maze passages (Ardoin, Witt, Suldo, Connell, Koenig, Resetar, Slider, and Williams, 2004) and the sequence for instruction of computation skills deviates little from one curriculum to the next (Hintze, Christ and Keller, 2002; Howell, Zucker and Moreland, 2000; Shapiro, 1996). Given that instruction in schools is increasingly tied to state content standards, and given that those standards in math include application and problem-solving skills, one might question if high achievement and mastery performance on assessments containing exclusively math computation problems would accurately reflect achievement and performance in the application and problem solving activities demanded in the classroom.

To begin to address this issue, a review of math curriculum-based assessment was undertaken. As stated above, of the three models of curriculum-based assessment described here, only CBM has a large body of research supporting the technical adequacy of the model and therefore the following review will focus on CBM.

Math Curriculum-Based Measurement (M-CBM)

Initially the research in this area focused on concurrent validity, presenting modest correlations between M-CBM and norm-referenced math tests or subtests. For example, early studies reported correlations lower than r = .60 with a median correlation of r =.425 between the Metropolitan Achievement Test (MAT) and first and sixth grade-level assessments (Skiba, Magnusson, Martson, and Erickson, 1986), first and second grade-level assessments (Skiba et al., 1986), third and fourth grade-level assessments (Skiba et al., 1986) and fifth and sixth grade-level assessments (Skiba et al., 1986). Other studies report similar findings. Martson (1989) and Putnam (1989) reported a median correlation of r = .43 between the MAT and problem solving skills and a r = .54 correlation between the MAT and computation skills (not an exhaustive review).

Following the initial validity studies, M-CBM research branched off in other directions. Fuchs, Fuchs, Hamlett and Stecker (1990), investigated the role of performance indicators and skills analysis in conjunction with M-CBM administration. Called the primary CBM datum, the performance indicator is an index of proficiency on the target skill. The most common performance indicator in mathematics is the number of digits correct per two minutes on a assessment that consists of sample items which represent the scope of the curriculum a student is expected to know at each grade-level, though again, this only pertains to computation problems. Therefore, early in the year, it is expected that the number of digits correct will be low compared to the number of digits correct at the end of the year. Progress is monitored by the performance indicator (i.e., math assessment scores) throughout the year using alternate forms of the assessment. The performance indicators are graphed and thereby enable teachers to a) monitor, adjust or proceed with student goals, b) determine if expected growth, as demonstrated by a trend line, will enable the student to obtain the year-end goal or if modifications are necessary, and c) assess intervention efficacy in a manner that is sensitive to intervention modifications. Fuchs et al. (1990) and others (Fuchs, Fuchs and Hamlett, 1989c; Fuchs, Fuchs, Hamlett and Stecker, 1991; Allinder, 1996) suggested that CBM used in this manner would differentially alter teacher behavior such that modifications would be made to instructional programming that would increase student-learning rates.

The secondary CBM datum, skills analysis, informs the teacher of the specific curricular skills not yet mastered (Fuchs et al., 1990). In their investigation, the authors suggested that skills analysis might be one potential strategy for enhancing CBM usefulness in curricular modifications in general and special education program development. Additionally, the authors suggested that the results of this study would contribute to the general field of assessment because test publishers are currently supplying skills analyses for the direct purpose of enhancing instructional usefulness of norm-referenced assessments. However, at the time of publication, researchers had not yet assessed the contributions of skills analysis to program development. Fuchs et al. compared two types of CBM models. The first model was CBM only. That is, CBM with only performance indicators guiding assessment and teacher made curriculum modifications. The second model incorporated CBM (with standard performance indicators) and skills analysis. The results suggested that skills analysis used in conjunction with CBM performance indicators enhanced teacher planning. That is, teachers using CBM with skills analysis were better prepared and constructed specific lesson plans based on the skills analysis results. Teachers using CBM without the skills analysis were not as prepared and did not have the same level of specificity in their lesson plans. Consequently, students in the skills analysis group performed better than students in the CBM performance indicator only group. The authors also concluded that skills analysis was an important factor in assessment and helped guide program development.

Allinder and Beckbest, (1995) investigated M-CBM and the type of follow-up support and the differential effects on teacher implementation and student growth. In the investigation, the authors compared the type of follow-up consultation offered to 18 teachers, 10 in the self-monitoring group and eight in the follow-up consultation group. Both groups received a two hour in-service training. Following in-service training, the self-monitoring group received a letter from the researchers describing the CBM self-monitoring procedures and 12 self-monitoring questionnaires to serve as permanent products of CBM implementation. The other eight teachers were assigned to the "university-based consultation" group. This group received bi-monthly consultation from graduate students. The graduate students provided answers to teacher questions, feedback on implementation and offered technical assistance. Results showed that there was no significant difference between the two groups of teachers (i.e., self-monitoring group or consultation group) regarding their use of the CBM data to modify individual instruction based on student need. Teachers in both groups suggested that one element they liked about the CBM assessments was that students were able to see their progress and this feedback to the students helped to motivate them and perform at a level closer to optimum. The researchers concluded that more research was needed regarding procedures that will assist or motivate teachers to use the data to make informed decisions regarding student instruction.

Subsequently, Allinder and Oats (1997) investigated the effects of teacher acceptability of CBM procedural administration. Based on the work of Witt and Elliot (1985), the authors hypothesized that teachers who like CBM procedures would be more inclined to implement CBM procedures. The results illustrated that teachers who reported liking CBM, implemented some of the procedures with greater fidelity than those who reported their dislike of CBM.

More recently, questions regarding the number of M-CBM assessments given during a single administration have been investigated. In other words, how many M-CBM assessments are needed to get an accurate assessment of student performance in that content area? Hintze, Christ and Keller (2002) questioned the necessity of multiple CBM assessments, as prescribed by the early literature, when determining performance indicators in mathematics. The authors stated that a particular single-skill mathematics assessment (e.g., addition sums to 18) varies slightly across forms and the algorithm used to solve the problems remains the same throughout the assessments. They argued that multiple skill math assessments, reading assessments, spelling assessments and writing assessments vary significantly in their level of difficulty and require the median of at least three assessments to obtain an accurate performance measure. Research by Hintze, Owen, Shapiro, and Daly (2000) supports this argument in reading. Therefore Hintze et al. examined the results of three single-skill and multiple skill computation assessments administered in grades one through five to determine if any practical difference was observed across the three scores. Results of repeated measures 3x5 (assessment by grade) ANOVA indicated no significant difference between the three single skill assessments within grades. The

3x5 repeated measures ANOVA for the multiple skill math assessments yielded different results. That is, significant differences were found in digits correct across grades one, two and five on the multiple skill assessments. Additionally, generalizability results indicated that performance on single skill math assessments cannot be generalized to performance on multiple skill math assessments and vice versa. Generalizability analysis within skill (i.e., single skill only or multiple skill only) indicates that performance on a single skill assessment can be generalized to overall performance on that single skill. Furthermore the single skill assessment is dependable for criterion and norm-referenced decisions. Dependability was slightly different on the multiple skill math assessment generalizability analysis. That is, the administration of one multiple skill math assessment is dependable for making norm-referenced decisions, but criterion-referenced decisions did not yield the same measure of dependability. They suggested that three assessments be given and the median score used as the performance indicator when making criterion-referenced decisions with multiple skill assessments. The authors concluded that single skill math assessments are dependable for instructional decision making and generalizable to class-wide or district norms. Multiple skill assessments were not as dependable for making changes to instructional planning. The authors suggested this is not unexpected given that the construction of a multiple skill assessment can vary significantly across assessments and therefore one assessment might yield few errors and therefore indicate that few modifications to instruction are necessary while the next assessment might yield many errors indicating considerable modifications to instruction are necessary. Generalizability results indicated that the single skill assessment results were generalizable to other single skill assessments. Similar results were not obtained for multiple skill probes and therefore and therefore taking the median of multiple assessments is still recommended.

A fundamental question regarding the validity of M-CBM is the extent to which this type of assessment is a valid indicator of overall performance in math. Thurber, Shinn and Smolkowski (2002) suggested two possible hypotheses regarding the historically low M-CBM validity coefficients with norm-referenced assessments. First, the authors suggested that the M-CBM lacked the same content found on the norm-referenced tests (e.g., state tests, achievement and IQ tests). Second, criterion measures include a significant amount of reading not found in the M-CBM assessments. In their investigation, Thurber et al., evaluated the two broad math constructs that achievement tests are traditionally based upon: operations (i.e., computation) and applications (i.e., problem solving). The Thurber et al. study follows from the work of Howell, Fox and Morehead (1993) and Silbert, Carnine and Stein (1990), who suggested that to solve math operation problems students must know the foundational strategies, concepts and facts. Whereas application problems require students to use and understand the concepts needed to solve word, measurement, volume, temperature, type problems (Salvia and Ysseldyke, 1991). Thurber et al. argued that M-CBM was developed to measure general math ability, not just computation. The authors suggested the theory that computation is a measure of general math ability is predicated on the purported relationship between computation and applications. Their investigation evaluated the relationship between M-CBM and the constructs of computation and applications as expressed on standardized assessments using confirmatory factor analyses. The study included 207 fourth grade students. The measures included three, fourth grade-level M-CBM assessments. The M-CBM assessments were constructed by sampling the annual curriculum of typical math texts. They ranged from simple basic math facts to complex problems requiring algorithms and strategies. However, it is important to note that the more complex problems were still computation problems and test content did not include application problems. The M-CBM assessments were five minutes in length and the number of digits correct determined the student's score. Students were also given two basic skill math assessments containing approximately 25% addition and subtraction problems and 75% multiplication and division problems. The multiplication and division facts consisted of exclusively single digit facts (e.g., 9 x 9, and 81 ÷ 9). Three maze CBM reading passages were included to evaluate the relationship between reading CBM and general math skills. The validity of the CBM measures were evaluated in conjunction with two norm-referenced tests, the Stanford Diagnostic Mathematics Test (SDMT) and the California Achievement Test (CAT). An additional assessment included application items from the National Assessment of Educational Progress (NAEP).

The results showed good alternate form reliability between the three CBM assessments (i.e., M-CBM, basic skill assessments and maze assessments). As expected, the M-CBM assessments correlated highest with the basic fact assessments, median correlation .82, but much lower with the application measures from the SDMT, NAEP and CAT with a median correlation of .44. The maze reading passages correlated highly with computation and applications, .76 and .77 respectively. The models derived from confirmatory factor analysis favored a two-factor model (rather than a single factor) where applications and computation problems were two separate, but related constructs. That is, the results suggest that computation and applications problems are relatively independent. Additionally, the reading assessments yielded high correlation coefficients with the applications assessments of the criterion assessments. Therefore, the authors suggested that reading proficiency may be necessary for a student to perform proficiently on a criterion-referenced assessment. One potential design limitation was that only computational skills were chosen. That is, the researchers did not develop applications problems from the local curriculum to compare to the criterion assessments.

Of the many M-CBM investigations of validity and sensitivity to instruction, the only study found investigating applications problems developed by the researchers and taken from the district curricula was conducted by Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek and Stecker (1994). In this study, the researchers discussed the importance of higher level problem solving skills in math and discuss the reduction in American competitiveness regarding doctorates earned in mathematics within American educational institutions. Additionally, the authors suggested that as the national economy and market moves towards jobs that require highly technical skills, including math skills, our students are becoming less proficient in problem solving skills. The authors suggested that M-CBM is a promising method for helping students learn math and directing teachers towards effective decision making regarding student instruction. As stated above, the authors noted that M-CBM assessment is limited to those items defined as "computation" problems, and therefore only address a limited sample of any math curricula. Therefore the authors constructed M-CBM assessments that incorporated application problems and were administered in accordance with traditional CBM methodology (i.e., weekly measures, with ongoing performance feedback to staff and students). The participants were 140 students in second, third and fourth grade classrooms. Within each of the classrooms (two per grade) were 1-3 students identified with learning disabilities, the rest were regular education students. The measures included two domains (applications and computation). The computation problems were grade-level basic facts similar to those assessments used in studies described above. The applications problems were constructed using the following procedures. First, the researchers analyzed the Tennessee math curricula standards. Next, the researchers assigned all math problems, other than computation, to a domain. Then they selected items from each domain from the curricula. Next, the researchers assigned a weight to each item in the domains and an item bank was developed. Following that, 30 alternate forms were constructed containing items from each domain. Then the researchers piloted the forms to identify potentially weak items. Finally, the assessments were administered weekly and student performance data was analyzed using two methods: 1) graphed performance on the assessments over time and 2) skills profiles showing student mastery of each domain. The domains identified in this process were; name numbers and vocabulary, measurement, charts and graphs, grid reading, areas and perimeters, fractions, decimals and word problems. Additionally, the researchers administered the Comprehensive Test of Basic Skills, fourth edition (CTBS) to use as a criterion measure. The CTBS was a district-wide spring assessment of student performance in mathematics on a nationally normalized “high stakes” assessment. The CTBS includes computation and applications problems.

Validity coefficients suggested that the applications assessment developed had moderate to high correlation coefficients with the CTBS (.64 to .81). Additionally, similar student growth in computation and applications problems were observed across the year for all grades ranging from $r = .40$ in second grade on applications to $r = .69$ in fourth grade and from $r = .25$ in second grade in computation to $r = .70$ in fourth grade. The authors concluded that assessing math applications could be done in combination with assessing computation. Some limitations not identified by the researchers however, are in order. First, concurrent validity was assessed with one only one assessment the CTBS. Therefore, conclusive statements are not yet in order. Second, the researchers do not compare the outcomes of the computation assessment with the CTBS, but rather compare the outcomes of the application assessments with the computation assessments only. As such, we do not know if success in applications problems is a better predictor of student performance in overall mathematics than computation problems. Third, the authors only sampled the Tennessee curriculum when reviewing the content standards used in the applications assessment construction and therefore some standards considered important in other regions of the country may be omitted. Minor limitations include an analysis of the applications assessment face validity, and concurrent validity with student grades and teacher report of student performance in applications and computation. This last limitation may be important when determining if teacher report is reliable when recommendations for special education are made.

Present Investigation

The present investigation was conducted to extend the findings of Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek and Stecker (1994). Given the paucity of literature that has evaluated CBM type assessments containing content beyond computation, the goal of the present investigation was to develop a procedure for constructing an math applications assessment which would have respectable concurrent validity with locally used criterion assessments. Furthermore, the present research addressed the question of whether a M-CBA type applications assessment have improved validity coefficients with normative criterion assessments than a single skill or multiple skill M-CBM assessment. Additionally, the study sought to answer questions pertaining to the face validity of various assessments in that successful use of an assessment is partially determined by whether teachers accept the scores derived. For example, many teachers report that the computation assessments are poor measures of student performance because they only measure computation. One question then is whether teachers prefer an assessment that includes applications problems in conjunction with computation problems.

Method

Participants and Setting

Participants included 173 regular education students from each of grades first through fifth. There were 19 males and 20 females in first grade, 16 males and 12 females in second grade, 13 males and 16 females in third grade, 24 males and 18 females in fourth grade, and 16 males and 19 females in fifth grade. All students attended a small urban public elementary school in southern Louisiana. The school ethnic makeup was approximately 46% African American, 46% Caucasian, and 7% Hispanic, 1% Middle Eastern and 1% Asian. Approximately 62% of students received a free or reduced lunch. Assessments took place during the last month of the school year, and approximately one and one half months after the administration of the Iowa Test of Basic Skills and Louisiana Educational Assessment Program for the 21st Century (see below). All assessments took place in the students' homerooms.

Measures

The present study examined the relationships between various assessments including (a) three math curriculum-based assessments (M-CBAs), (b) one reading curriculum-based assessment, and (c) three norm-referenced assessments. In addition teacher report data were collected on teacher preference of M-CBAs, students' final grades and year-end student performance in math applications and computation problems. These measures are described below.

Math – Curriculum Based Assessment: Applications and Computation

The items for this assessment included the two main constructs (i.e., applications and computation) of general mathematics described by Thurber et al., (2002). To begin the development of this new math curriculum-based assessment, we randomly selected and analyzed mathematics curricula standards from seven states across the country (Louisiana, South Carolina, Arizona, Alaska, Connecticut, Mississippi, and Massachusetts) to determine the types of application problems (i.e., problem solving problems) included in a sample of the nations schools. Although multiple standards were noted for each state, there was a common core of six standards listed by five states; a) number sense, b) data analysis, c) patterns, d) algebra, e) geometry and f) measurement, and five standards listed by all seven (see Appendix A for a definition of each area). These content standards also coincide with the content strands for mathematics prescribed by the National Assessment of Educational Progress (NAEP) also known as the "Nation's Report Card". Additionally, Louisiana, the state in which the study took place, identified all six standards as the core math content for the state (Appendix A).

Next, we obtained a copy of the math curriculum used by the participating district (*Math Central*: Houghton Mifflin, 2003). Additionally, district content standards, which mirrored the six Louisiana state standards, were obtained. Finally, a curriculum guide that accompanied Math Central was obtained that specified which items on the chapter tests of the math curriculum aligned with the six content standards for the district and state. A table of specifications for each grade was then constructed such that the six content standards were listed in columns. Next, the grade-level chapter tests were reviewed and items were identified by content standard according to the district table and listed in the table of specifications under the designated standard. Once the items were identified and listed in the table of specifications, I determined if they fit the definition of each content standard (e.g." algebra") as defined by the states' content standards and the NAEP (Appendix A). Then six items were randomly selected, and modified (to protect the publishers copyright) from each standard for inclusion on each grade-level assessment with one exception. The items were randomly selected from the chapter tests beginning with early chapters through to the later chapters (rather than randomly from all chapters). Hence, item difficulty increased in a sequential fashion across the assessment. Therefore, each grade-level assessment contained 36 total items.

To insure the accuracy of this process, a reliability check was invoked. The check was conducted by a content expert who was a district Special Education Administrator with a Doctorate in Education. The content expert was responsible, in part, for district curricula selection. The content expert reviewed district/state standards and analyzed the modified problems for errors in operation and/or answer. If an error was identified the problem was either corrected or discarded. The process continued until there was 100% agreement that each grade-level item reflected the designated standard.

An example will clarify how the process for item selection operated. First, the district provided a table that identified which problems represented one of the six content standards in each grades chapter tests. I turned to the page number of the first chapter that listed algebra problems and found the item identified in the district table. That search would produce a problem such as this: “Write >, <, or = in the following problem. 298 _____ 289.” I entered the problem into the table of specifications and continued to identify items from the remaining chapters. That process continued until one algebra problem was selected from each chapter of those containing algebra problems. Next, the number of items representing algebra was reduced to six. Hence, if 12 algebra problems were listed in the table of specifications, then we randomly choose one problem from the first two listed, one from the second two listed and so on until we had six problems. The problems were then modified slightly so that the problem above might appear as: “Write >, <, or = in the following problem. 278 _____ 269.” The content expert then reviewed the items and withdrew any items for inaccuracies. This process occurred for all grade-level assessments. (See Appendix B for completed grade-level assessments.)

Math – Curriculum Based Assessment: Single Skill Computation

The single skill computation assessment was comprised of one basic math skill (e.g., addition, subtraction, multiplication or division). To determine grade specific and content relevant material, teachers were asked to nominate a "recently taught" basic skill on which most students were expected to perform proficiently. Skills nominated where then compared to 2003 district math curriculum standards to insure that computational skills chosen by the teachers were those prescribed by the district. First grade teachers chose addition sums to 14, second grade teachers chose addition sums to 18, third grade teachers chose multiplication facts to seven, fourth grade teachers chose multiplication facts to nine, and fifth grade teachers chose division from 81-0 (actual assessments will be provided upon request).

Once the content areas were selected, assessments were generated using software available from *Schoolhouse Technologies* called Mathematics Worksheet Factory Deluxe V3 © (1998-2002). The software allows users to create the computation worksheets (i.e., assessments). A feature called a worksheet generator was used to create the worksheets.

The worksheet generator used for the basic skills assessments was called Number Operations (Appendix C). The worksheet generator has actions available to adjust a number of variables including problem difficulty (e.g., sums to five, sums to 18, multiplication times tables 1-20), number of problems per page, and problem orientation (i.e., horizontal or vertical presentation). A trial version of the software can be found at http://www.schoolhousetechnologies.com.

Math – Curriculum Based Assessment: Multiple Skill Computation

To select the content for the multiple skill computation assessment, district curriculum standards were reviewed. The results of that review were then discussed with each participating teacher to insure that those specific skills had in fact, been taught during the year. All teachers agreed that the skills chosen for these assessments were taught during the year. First grade multiple skill assessments included the skill addition sums to 14 and subtraction from 14. Second grade included addition sums to 18 and subtraction from 14. Third grade included addition sums to 99, subtraction from 99 and multiplication facts to and including five. Fourth grade items included addition sums to 99, subtraction from 99 and multiplication facts to and including nine and division from nine. Fifth grade included addition sums to 1000, subtraction from 1000, multiplication facts to and including nine and division from nine.

The multiple skill assessments were developed using the same software, worksheet generator and process as described for the construction of the single skill assessment with this exception: the multiple skill assessment contained two or more basic math skills.

Reading – Curriculum-Based Assessment: Maze Reading Passages

The content controlled passages were selected from a reading intervention created by the School Psychology Department at Louisiana State University. The reading intervention contained a set of 36 passages (12 sets of three alternate-forms) for each grade one through five. The passages chosen for this research were entered into the software Readability Calculations v. 3.7 © (2000), *from Micro Power and Light Company* to ensure grade-level readability. Two formulas were used to determine passage readability. The Spache readability formula, developed for early elementary grades, was used for grades one through four.

The Dale-Chall readability formula, which can be used for middle school and high school, was used for the fifth grade passages. Information about the reading passages can be found at http://bitwww1.psyc.lsu.edu/reading%20center.htm.

Once the three alternate forms were checked for grade-level readability, they were converted into maze reading passages using the procedure described by Shinn (1998). That is, every seventh word was deleted from the content controlled reading passages and replaced with a blank line except for the first sentence, which remained intact. To the right of each blank were three words in parentheses from which the student could choose. One of the three words correctly completed that portion of the sentence and the other two words had no relation to the sentence. The students identified their choice by circling one of the three words.

Teacher Report Measures

<u>Teacher rating of student performance and expected final grade.</u> We asked each teacher to assess their students' overall performance in math in the two broad areas described above (i.e., applications and computation). That is, teachers were asked to categorize their students' math competence in applications and computation according to one of the following six levels of a Likert scale; a) mastery, b) instructional, c) satisfactory with help, d) some difficulty e) frustrational and f) cannot perform this skill (Appendix D). Teachers were also asked to indicate the students' final grade.

<u>Teacher preference of math curriculum-based assessments.</u> Math curriculum-based assessment face validity was determined by modifying the Intervention Rating Profile 15 (Witt and Martens, 1984). Modifications included substituting all references to "Intervention" with "Assessment" and rephrasing questions such that they asked about the preference of the

assessment, rather than the intervention. (Appendix E). A comparison of the two rating profiles will be made available upon request.

Iowa Test of Basic Skills, (ITBS)

The ITBS is a national norm-referenced achievement assessment published by *Riverside Publishing* used to compare local students' scores with students across the country. Of the grades we assessed, Louisiana offers the ITBS to grades two, three and five. Louisiana does not offer the ITBS to first grade, and fourth grade is administered a criterion-referenced assessment described below. The ITBS is one of many standardized assessments used nationally by education agencies to assess and report student achievement in academics. Content areas covered in batteries seven and higher (those administered to grades two, three and five in Louisiana) include the same content standards identified by the state content standards sampled (e.g., number sense, operations, measurement, geometry, data analysis, algebra and patterns). The ITBS can be individually or group administered and is typically administered in the spring.

The purpose of the ITBS is to report students' academic strengths and weaknesses and compare students' scores within grade and across districts and states. Based on Kuder-Richardson Formula 20 (KR-20) procedures, ITBS test reliability estimate for math computation, problem-solving and data interpretation, and math concepts and estimation range from .761 to .906 for grades one through five.

The Northwest Evaluation Association (NWEA) obtains validity data for many of the state administered, high-stakes norm-referenced achievement tests such as the ITBS. In order to obtain Pearson correlation coefficients and determine concurrent validity, the NWEA compares the results norm-referenced achievement test (i.e., ITBS) to either other norm-referenced assessments or student measures of academic performance. For the ITBS, the NWEA compared the results of the Meridian Checkpoint Levels Tests with that of the ITBS. Correlation coefficients exceeded .7 (third grade) and .8 (fifth grade) in mathematics (Northwest Evaluation Association, 2004).

Louisiana Education Assessment Program for the 21st Century (LEAP 21)

LEAP 21 is the state sanctioned criterion-referenced assessments used to monitor how well students have learned the state content standards up to fourth and eighth grades. The purpose of LEAP 21 was therefore, to ensure that fourth and eighth grade students had the knowledge to pass onto the next level of education (i.e., middle school and high school). To determine this, students are given achievement ratings; advanced, mastery, basic, approaching basic and unsatisfactory. In order for students to move to the next grade, they must have and achievement rating of approaching basic or above.

The assessments were developed using items developed by testing contractors and approved of by an advisory committee. The items were judged on congruence with the state standards specifications, technical quality and age-appropriate content validity. Additionally, a Bias Review Committee critiqued the items for gender, ethnicity and special population issues. The assessments were then field tested in randomly selected schools based on the following stratifications: size, ethnicity, socio-economic status, and achievement performance. The data from the field trial was then submitted to the advisory committee for final review. The committee then determined based on technical adequacy which items were to be retained for the assessment.

Woodcock-Johnson Psychoeducational Battery III, (WJ III)

According to manufacturers, The Woodcock-Johnson Psychoeducational Battery III is a norm-referenced, comprehensive system that can be used to measure a student's academic achievement, general intellectual ability, oral language, scholastic aptitude and cognitive abilities. The WJ has two parts, the Cognitive Battery and the Achievement Battery. The Achievement Battery is broken into two broad categories of language arts and mathematics, and parallels the Individuals with Disabilities in Education Act, 1997 areas for determining discrepancies between a student's achievement level and his ability.

The WJ III has test-retest reliability in math fluency of .75 in the 7-10 age range and .86 in the 8-11 age range, a test-retest reliability in applied problems of .92 in the 7-10 age range and .85 in the 8-11 age range and a test-retest reliability in calculation of .87 in the 7-10 age range and .83 in the 8-11 age range.

The WJ III normalization sample was stratified, within practical constraints, from the United States population from ages 24 months to 90+ years. Test validity depended on two factors: 1) the representation of the sample to the population and 2) the careful data collection from the sample. Ten individual and community factors and 13 socio-economic factors were taken into consideration when deciding sample participants.

For the purposes of this investigation, we were only interested in the correlations between the WJ III mathematics subtests scores, and the M-CBAs, therefore only the three math subtests (i.e., calculation, fluency and applied problems) were included.

Procedure

The administration of the measures included in this study was timed to follow the administration of the various required accountability testing for the district. Hence, the first assessments administered were the Iowa Test of Basic Skills and Louisiana Education Assessment Program for the 21st Century. These measures were conducted exclusively by district professionals. Six weeks following the ITBS and LEAP administration, we administered the math and reading curriculum-based assessments and the WJ III math subtests.

Training Data Collectors

Psychology doctoral students attending a university in Louisiana were trained in the administration of curriculum-based assessment (CBA) and Woodcock-Johnson III (WJ III) procedures. The math and reading curriculum-based assessment procedures were similar to those described in Shinn (1989) and are described below. Graduate students were required to memorize the procedures and re-state the procedures upon request without assistance. Graduate students were required to read the written procedures during the CBAs with 100% accuracy (Appendix F). Additionally, graduate students were trained on the WJ III administration procedures described in the manual for the math subtests and required to be 100% accurate with the written procedures before administering the subtests with students.

Interrater Agreement and Procedural Integrity

Interrater agreement was determined by having a second person score approximately 20% of the math CBA assessments, Maze reading passages and WJ III math protocols. ITBS and LEAP agreement was determined by having a second graduate student randomly check the values reported by the testing agencies, with the values collected by the primary investigator. That is, once the testing agencies returned the students' scores, approximately two months after the assessments were given; the primary investigator recorded the values with the corresponding student names into an electronic spreadsheet. The second graduate student was then given an electronic copy of the spreadsheet for the integrity checks with the values reported in the testing agencies report. The district held all ITBS and LEAP reports in a sealed office in the district administration building.

Procedural integrity was determined by having an independent observer observe approximately 20% of the math and reading CBAs and WJ III battery administrations and indicate whether all, some or none of the procedural steps were completed (Appendix G).

Assessment Administration

The curriculum-based assessments were administered to all students in all grades participating in this study. The ITBS was administered by the state to grades two, three and five and the LEAP was administered by the state to grade four (as described above). The math subtests of Woodcock-Johnson Psychoeducational Battery III (WJ-III) was randomly administered to one class (approximately 15 – 20 students) in each grade so as to allow the data collection for this instrument to be completed within one week of beginning the curriculum-based assessments. Teacher report data were collected on all students participating in the study when assessment data collection was complete.

Iowa Test of Basic Skills administration, (ITBS). Classroom teachers, school counselors and other ancillary school personnel administered the ITBS according to the prescribed procedures that accompany the test. Tests were administered and then returned to the publisher for scoring. Results were sent back to the district approximately eight to ten weeks later.

Louisiana Education Assessment Program for the 21st Century (LEAP 21). Classroom teachers, counselors, and school personnel administered the LEAP according to the standardized procedures that accompany the test. Scoring was conducted by the test publishers and the results were returned to the district approximately six to eight weeks later.

Math – curriculum based assessment: Single skill computation. After the teacher identified the basic skill and level of problem difficulty, one assessment, as specified by Hintze, Christ and Keller (2002) was generated using the procedures described above. Modified procedures similar to those described by Shinn (1989) were used to administer the single skill math assessment (Appendix F). That is, a class-wide administration was employed whereby assessments in the form of single page worksheets were distributed to all students in the class. Students were requested to keep the worksheets face down on their desks and write their names and their teachers name on the back of the worksheet. The students were informed of the type of problem on the worksheet (e.g., addition, subtraction, multiplication or division) and told they had two-minutes to complete as many problems as they could. Students were then asked to start at the top left of the page and to work across the page to the right, then to move on to the next row. Next, students were instructed not skip around and not to leave any answers blank. Finally, students were asked if they had any questions, then instructed to begin the assessment. An electronic timer was used to count-down from two-minutes. When the timers counted down to zero, a loud audible sound was made and students were asked to stop working.

Math – curriculum based assessment: Multiple skill computation. Using the math curriculum-based assessment development procedures described above, three alternate parallel forms were generated as suggested by Shinn (1989) and Hintze, Christ and Keller (2002). Assessment administration was the same single skill computation procedure described above (Appendix F).

Math - curriculum-based assessment: Applications and computation. The applications and computation assessment had similar administration instructions as the two previous math assessments described above. Hence, students were asked to keep their assessments face down until the assessment began, they were asked to write their name and their teachers' name on the back of the assessment, and the were instructed that they would find problems that assessed their knowledge in number sense, algebra, patterns, data analysis, geometry and measurement.

However, the procedure differed in that we were interested in both fluency and accuracy measures (Shinn, 1989). Fluency was addressed by allowing students five minutes to complete as many problems as possible. This is similar to the two minutes afforded students in the single skill and multiple skill computation assessment. After the five minute fluency segment, a graduate student circled all attempted or completed problems for later scoring (described below in the Data Collection and Scoring section). Immediately following the fluency measure, the students were asked to complete the assessment and instructed to hold the assessment over their head for collection when they were finished. This segment was completed to obtain an accuracy measure.

Reading – Curriculum-based measurement: Maze reading passages. In accordance with procedures described by Shinn (1998), three alternate-forms Maze reading passages were administered class-wide (Appendix F). That is, one passage was put on each student's desk in the classroom. The student's were then asked to turn the passage over when instructed, to read the passage up to the blank in the sentence and then to determine which word best finished the sentence. The students were instructed to circle the word and continue with the passage. They were told they had two-minutes to complete as many sentences as possible. The procedure was then repeated two times.

Woodcock-Johnson Psychoeducational Battery III, (WJ III). Three psychology doctoral students administered the math subtests of the WJ III according to the procedures described in the administration manual. Administration occurred during the same week the curriculum-based assessments were being administered.

Data Collection and Scoring

The LEAP and ITBS were scored by the manufacturers. The results were then shown to the primary investigator by the district six to ten weeks after the assessments. The primary investigator then entered the math scores for the ITBS and LEAP assessments into an electronic spreadsheet. A second investigator then randomly checked the values in the spreadsheet against the values for each students' math scores on the ITBS and LEAP provided by the district.

<u>Math - curriculum-based assessment: Single and multiple skill.</u> Math CBAs were scored by the graduate students in accordance with the procedures derived form Shinn (1989 and 1998). Single skill and multiple skill math CBAs were scored by tallying the number of digits correct in a two-minute assessment. Digits correct were defined as those numeric values placed in the correct sum, difference, product and quotient columns (e.g., ones, tens, and hundreds). Partially completed problems were included if part of the answer was correct and in the right column. Correct digits also included correct values as remainders in division problems, and numbers used for regrouping in subtraction and carrying in addition.

<u>Math - curriculum-based assessment: Applications and computation.</u> The math applications and computation assessment was scored using two procedures; 1) digits correct, and 2) total problems correct for both *fluency* and *accuracy* measures. Therefore four measures were obtained; 1) fluency digits correct (FLDC), 2) accuracy digits correct (ACDC), 3) fluency total correct (FLTC) and 4) accuracy total correct (ACTC). Digits correct were determined using the same procedure described above with these exceptions. First, the fluency assessment was increased to five-minutes to allow for the additional time needed to read the applications problems. Second, digits correct were also tallied for all problems completed, not just those completed in five-minutes. Hence, digits correct were tallied for the fluency measure (i.e., five-

minute assessment) and an accuracy measure (i.e., all assessment problems). Total problems correct were also tallied for both the five-minute fluency assessment and the total number of correct problems for the assessment. The scoring of problems correct differed from the scoring for digits correct in that the whole answer needed to be 100% correct to be added counted.

Curriculum-based assessment: Maze reading passages. Maze passages were scored by tallying the number of correct choices a student made in a two-minute timed assessment. A student made a correct choice when they circled the one word of three that correctly completed the sentence.

Results

This research focused on the extent to which scores on the newly developed math assessments in applications and computation were associated with performance on the single and multiple skill computation assessments and standardized criterion assessments (i.e., ITBS, WJ III and LEAP). As such, Pearson product-moment correlation coefficients were the primary methods of data analysis. Following the correlation analyses, the correlation coefficients between the criterion variable (i.e., LEAP, WJ III and ITBS) and the predictor variables (i.e., single skill and multiple skill computation, maze reading and applications assessment) were rank ordered, and a test of significance of the difference between dependent correlation coefficients (Glass and Stanley, 1970) was conducted. In only one case did a significant difference between correlation coefficients emerge. This occurred in the set of analyses for the fourth grade which revealed that the correlation coefficient for the ACTC ($r = .729$) was significantly higher than the correlation coefficient for the SSC ($r = .228$, $t(16) = 2.01$, $p < .05$). Given the large number of analyses performed, this significant difference between correlations may have occurred by chance.

For each grade a set of secondary analyses were performed. These analyses looked at the extent to which students' final grades, teacher report of student performance on computation and applications problems in class, and the math curriculum-based measurements correlated. Finally, we evaluated teacher preference for the math curriculum-based assessments as a way to measure the face validity of the assessments. Results are presented below by grade.

All correlations were computed using a family wise model. That is, the alpha level for the specific question asked was determined by taking the number of correlations and dividing them by .1. For example, if a specific question required ten correlations, then we divided .1 by ten to get an alpha level of .01. Thus, for that analysis, only .01 correlations were determined to be statistically significant. Additionally, data reported in the tables may not have been analyzed in the same family, but are presented together to allow for comparisons. Finally, all tables reporting the primary and secondary analyses for each grade present correlation coefficients for the same set of variables regardless of significance so that the reader may review all the findings and compare the results across the grades.

As described above, four scoring procedures were used to score the newly created math curriculum-based assessments in applications and computation (see Method section above). Initial analyses revealed that the five-minute fluency measures of total problems correct (FLTC) and total digits correct (FLDC) produced few significant results at the .01 level and, and therefore to simplify data presentation, those data are not reported here (data will be made available upon request). However, scoring the measures for accuracy (see Method section above) of total problems correct (ACTC) and total digits correct (ACDC) produced scores which routinely yielded statistically significant relationships with other variables and therefore, were reported for all grade-level comparisons.

Data Reported and Table Designations

Woodcock-Johnson III math measures included broad score (BS), calculations (CALC), fluency (FL) and applied problems (AP). ITBS measures included ITBS composite score (IC), total math score (TM), math composite score (MC), and total reading score (TR). The Louisianan Education Assessment Program measure for math was LEAP total math (LTM). Math curriculum-based assessment measures included single skill computation (SSC), multiple skill computation (MSC) and the new applications and computation (AC) assessment. The AC measures reported (as stated above) were: 1) accuracy - total problems correct (ACTC), and 2) accuracy - total digits correct (ACDC). Maze reading assessments were designated (MZR).

The data presentation for each grade are organized such that the correlations between the ACDC and ACTC are reported in the beginning of each grade. Additionally, MZR passages and MSC assessment correlations are reported in the beginning of each grade's results to illustrate the correlations across the multiple forms. The median MZR reading passage (Ardoin et. al. 2004) and MSC assessments (Hintze, and Christ, 2003) were used as the comparison score for all other curriculum-based assessment correlations with the standardized assessments reported in the tables.

Implementation Integrity and Observer Agreement

Procedural integrity for the reading and math curriculum-based assessments and WJ III math subtests was determined by having a second person observe the administration of the assessments and indicate on Appendix G all steps completed. Twenty-two percent of the assessments were observed and procedural integrity, defined as the total number of instructions given divided by the total number of instructions was 98%.

Inter-scorer agreement was determined by having a second person score the math and reading curriculum-based assessments (MZR, SSC, MSC, ACTC and ACDC). Twenty five percent of the assessments were scored and total score agreement was 92%. Total score agreement for the SSC and MSC were digits correct. Total score agreement for the ACTC and ACDC was total problems correct and total digits correct respectively. Total score for the MZR was total words correct. Additionally, 98% of the LEAP and ITBS values were checked by a second observer and compared to the values recorded in a spreadsheet by the primary investigator. Observer agreement was 98% for LEAP and ITBS values.

First Grade

Primary analyses. First, Pearson product-moment correlations were computed to obtain estimates of alternate form reliability for the MSC and MZR passages. The alternate forms of the MSC assessment were: assessments 1 and 2, $r = .835$ $p < .01$; assessments 1 and 3, $r = .768$ $p < .01$ and assessment 2 and 3, $r = .731$ $p < .01$. The MZR passage alternate form correlations were: forms 1 and 2, $r = .885$ $p < .01$; forms 1 and 3, $r = .583$ $p < .01$ and forms 2 and 3, $r = .696$ $p < .01$. The correlation between the ACTC and ACDC was $r = .986$ $p < .01$.

Next, a series of Pearson product-moment correlations were computed between the various predictor variables (e.g., CBM) and the various criterion variables (e.g., WJ III). Table 1 illustrates the correlations between the four WJ III math sub-scores and the math curriculum-based assessments. Slightly higher correlations were obtained between the WJII BS and ACTC, $r = .785$ $p < .01$ than the other predictor measures. Higher correlations were obtained for the ACTC and the WJII Fluency measure ($r = .793$ $p < .01$ and $r = .774$ $p < .01$) than the SSC and MSC assessments. No significant correlations were obtained for the WJII CALC measure. Correlations between the WJ AP and the MSC, $r = .781$ $p < .01$, were statistically significant and indicated a stronger association than for the AC predictor measures (ACTC and ACDC).

Table 1. First Grade Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level		Woodcock-Johnson III Math Measures			
		BS	Calc	FL	AP
Math Curriculum-based Measures	SSC	.708**	.461	.714**	.645**
	MSC	.760**	.409	.662	.781**
	ACTC	.785**	.619	.793**	.610
	ACDC	.735**	.590	.774**	.562

Secondary analyses. Next, Pearson product-moment correlations were computed and shown in Table 2 between students' final grades, math curriculum-based assessments (i.e., predictor variables) and standardized assessments (criterion variables). Highest correlations obtained were those between final grades and ACTC, r = .707 p< .01 and ACDC, r = .723 p< .01.

Table 2. First Grade Students' Final Grades, Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level	Woodcock-Johnson III Math Measures			
	BS	Calc	FL	AP
Students' Final Grades	.474	.475	.619**	.290
	Math Curriculum-based assessments			
	SSC	MSC	. ACTC	ACDC
Students' Final Grades	.568**	.578**	.707**	.723**

Table 3 illustrates the correlations between teacher report data regarding student performances on applications and computation problems as reported on the rating scale (Appendix C). The highest correlations were those between the final grades and teacher report of student performances on applications and computation problems.

Table 3. Teacher Report, Final Grade and Math Curriculum-Based Assessment Correlations for First Grade.

** sig. at .01 level		Math Curriculum-based assessments				
		Final Grade	SSC	MSC	ACTC	ACDC
Teacher Report	Computation	.912**	.588**	.638**	.721**	.733**
	Applications	.908**	.673**	.676**	.744**	.747**

Of the predictor variables, the ACTC and ACDC obtained the highest correlations with both teacher report of student performance on computation, r = .721 p< .01 and r = .733 p< .01 respectively; and applications and ACTC, r = .744 p< .01 and ACDC, r = .747 p< .01.

Second Grade

Primary analyses. The correlation between the ACTC and ACDC measures was r = .869 p< .01. The MSC assessment alternate form correlations were: assessments 1 and 2, r = .859 p< .01; assessments 1 and 3, r = .809 p< .01 and assessment 2 and 3, r = .835 p< .01. The MZR reading passage alternate form correlations were forms 1 and 2, r = .925 p< .01; forms 1 and 3, r = .902 p< .01 and forms 2 and 3, r = .856 p< .01.

Table 4 illustrates the correlations between the math curriculum-based assessments and the WJIII. Results show that the highest significant correlation at the .01 level between the WJII BS was with the AC measure ACTC, r = .754 p< .01. The ACTC and ACDC measures also had the highest correlations with the WJII AP measure, r = .812 p<.01 and r = .667 p<.01 respectively. Significant correlations between the math curriculum-based assessment probes and the WJII FL measure were observed for the SSC and MSC assessments, with the SSC assessment having the strongest degree of association, r = .780 p<.01.

Table 4. Second Grade Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level		Woodcock-Johnson III Math Measures			
		BS	Calc	FL	AP
Math Curriculum-based Measures	SSC	.434	.459	.780**	.253
	MSC	.669**	.620	.745**	.522
	ACTC	.754**	.694**	-.194	.812**
	ACDC	.544	.330	-.349	.667**

Table 5 illustrates the correlations between the ITBS, MZR reading and the math curriculum-based assessment measures.

Table 5. Second Grade ITBS Test of Basic Skills and Curriculum-Based Assessment Correlations.

** sig. at .01 level		ITBS Test of Basic Skills IC	TM	MC	TR
Math Curriculum-based Measures	SSC	.249	.446	.452	.350
	MSC	.338	.486	.613**	.483
	ACTC	.679**	.814**	.752**	.651**
	ACDC	.575**	.652**	.650**	.624**
Reading	MZR	.654**	.544**	.434	.755**

The ACTC and ACDC scores had higher correlations with the IC, $r = .679$ $p<.01$ and $r = .575$ $p<.01$ respectively; ITBS TM, $r = .814$ $p<.01$ and $r = .652$ $p<.01$ respectively; ITBS MC, $r = .752$ $p<.01$ and $r = .650$ $p<.01$ and the ITBS TR, $r = .651$ $p<.01$ and $r = .624$ $p<.01$ than the other math curriculum-based measurements. Additionally, scores on the MZR passages had the highest, significant correlations with the ITBS TR score, $r = .755$ $p<.01$. The SSC had no significant correlations with the ITBS scores. The MSC assessment had significant correlations with ITBS MC, $r = .613$ $p<.01$.

Secondary analyses. Table 6 shows the correlations between the standardized assessments (i.e., Woodcock-Johnson III and ITBS Test of Basic Skills), the math curriculum-based assessments and students' final grades. Final grades were not correlated with any Woodcock-Johnson III measures.

Table 6. Second Grade Students' Final Grades, ITBS Test of basic Skills, Woodcock-Johnson III and Curriculum-Based Assessment Correlations.

** sig. at .01 level	ITBS Test of Basic Skills			
	IC	TM	MC	TR
Students' Final Grades	.863**	.791**	.663**	.808**
	Woodcock-Johnson III Math Measures			
	BS	Calc	FL	AP
Students' Final Grades	.497	.304	-.306	.595
	Math Curriculum-based Assessments			
	SSC	MSC	. ACTC	ACDC
Students' Final Grades	.568**	.373	.707**	.723**

ITBS measures had high correlations with students' final grades, r = .863 p<.01 with IC, r = .791 p<.01 with ITBS TM, r = .633 p<.01 with ITBS MC and r = .808 p<.01 with ITBS TR. Of the math curriculum-based assessments, the highest correlations with students' final grades were those for the ACTC, r = .707 p<.01 and ACDC, r = .723 p<.01. Significant correlations were obtained for the SSC assessment, r = .568 p<.01.

Table 7 presents the correlations between teacher report measures and the math curriculum-based assessments. Significant correlations observed were for teacher report of student performances on applications and computation problems with the ACTC and computation r = .700 p<.01; ACTC and applications = r = .606 p<.01; ACDC and computation = r = .562 p<.01; and ACDC and applications = r = .628 p<.01.

Table 7. Teacher Report, Final Grade and Math Curriculum-Based Assessment Correlations for Second Grade.

** sig. at .01 level		Math Curriculum-based assessments				
		Final Grade	SSC	MSC	ACTC	ACDC
Teacher Report	Computation	.893**	.424	.432	.700**	.562**
	Applications	.787**	.207	.239	.606**	.628**

Third Grade

Primary analyses. The ACTC and ACDC correlation was, r = .987 p<.01. Alternate form coefficients for the MSC were as follows: 1 and 2, r = .842 p<.01; 1 and 3, r = .681 p<.01 and 2 and 3, r = .730 p<.01. Maze reading alternate form correlations were as follows: 1 and 2, r = .819 p<.01; 1 and 3, r = .744 p<.01 and 1 and 3 = r = .614 p<.01.

Table 8 illustrates the correlations obtained between the Woodcock-Johnson III subtests and the math curriculum-based assessments. Significant correlations were observed between the ACTC assessment and the WJ III BS, r = .701 p<.01, the MSC and the WJ III CALC, r = .730 p<.01, and the WJ III FL, r = .689 p<.01. The AC assessment had significant correlations with the WJ III AP score (ACTC and WJ III AP, r = .705 p<.01, and ACDC and WJ III AP, r = .693 p<.01).

Table 8. Third Grade Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level		Woodcock-Johnson III Math Measures			
		BS	Calc	FL	AP
Math Curriculum-based Measures	SSC	.447	.452	.470	.295
	MSC	.691**	.730**	.689**	.426
	ACTC	.701**	.621**	.441	.705**
	ACDC	.672**	.547	.470	.693**

Table 9 illustrates the correlations between the ITBS Test of Basic Skills and the curriculum-based assessment scores. Significant correlations were observed between both scoring methods for the applications and computation assessment (ACTC and ACDC) and all measures of the ITBS. Interestingly, SSC and MSC scores did not yield a significant correlation with TM but MZR did.

Table 9. Third Grade ITBS Test of Basic Skills and Curriculum-Based Assessment Correlations.

** sig. at .01 level		ITBS Test of Basic Skills			
		IC	TM	MC	TR
Math Curriculum-based Measures	SSC	.531**	.500	.466	.434
	MSC	.572**	.532	.544**	.439
	ACTC	.689**	.738**	.594**	.586**
	ACDC	.670**	.691**	.608**	.574**
Reading	MZR	.560**	.527**	.260	.520**

Secondary analyses. Table 10 illustrates the correlations between students' final grades, the standardized assessments and the math curriculum-based assessments.

Table 10. Third Grade Students' Final Grades, ITBS Test of basic Skills, Woodcock-Johnson III and Curriculum-Based Assessment Correlations.

** sig. at .01 level	ITBS Test of Basic Skills			
	IC	TM	MC	TR
Students' Final Grades	.805**	.756**	.563**	.711**
	Woodcock-Johnson III Math Measures			
	BS	Calc	FL	AP
Students' Final Grades	.644**	.510	.252	.808**
	Math Curriculum-based Assessments			
	SSC	MSC	. ACTC	ACDC
Students' Final Grades	.599**	.677**	.781**	.761**

Significant correlations between students' final grades and standardized measures were: final grades and ITBS Composite Score, r = .805 p<.01, and students' final grades and WJ III Applied Problems, r = .808 p<.01. Significant correlations between students' final grades and math curriculum-based assessments were obtained with all math CBA measures.

Table 11 illustrates the correlations between teacher report of student performances on applications and computation and math curriculum-based assessment measures. Results show that the ACTC score had the highest correlations with teacher report of student performance on both computation, r = .673 p<.01, and applications, r = .690 p<.01 but all correlations were statistically significant.

Table 11. Teacher Report, Final Grade and Math Curriculum-Based Assessment Correlations for Third Grade.

** sig. at .01 level		Math Curriculum-based assessments				
		Final Grade	SSC	MSC	ACTC	ACDC
Teacher Report	Computation	.916**	.573**	.599**	.673**	.657**
	Applications	.890**	.549**	.610**	.690**	.673**

Fourth Grade

Primary analyses. Correlation obtained between ACTC and ACDC was r = .914 p<.01. Multiple skill computation assessment alternate form coefficients were: 1 and 2, r = .796 p<.01; 1 and 3, r = .781 p<.01 and 2 and 3, r = .793 p<.01 . Maze reading passage alternate form correlations were: 1 and 2, r = .737 p<.01; 1 and 3, r = .869 p<.01 and 2 and 3, r = .878 p<.01.

Table 12 shows the Woodcock-Johnson subtest and math curriculum-based assessment correlations. Highest significant correlations were observed between applications and computation measures ACTC and ACDC with the WJ III BS: r = .699 p<.01 and r = .690 p<.01, respectively and the ACTC and WJ III AP: r = .613 p<.01. The WJ III FL score yielded the highest correlations with the SSC, r = .639 p<.01 and MSC, r = .634 p<.01.

Table 12. Fourth Grade Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level		Woodcock-Johnson III Math Measures			
		BS	Calc	FL	AP
Math Curriculum-based Measures	SSC	.233	.078	.639**	.030
	MSC	.569	.486	.634**	.300
	ACTC	.699**	.506	.205	.613**
	ACDC	.690**	.474	.382	.559

Table 13 illustrates the correlations between the curriculum-based assessments in reading and math and the Louisiana Education Assessment Program for the 21st Century (LEAP 21). The data indicates that significant correlations can be found between the LEAP 21 TM score and the ACTC, $r = .729$ $p<.01$, and the ACDC, $r = .691$ $p<.01$.

Table 13. Fourth Grade Louisiana Education Assessment Program for the 21st Century (LEAP 21) and Curriculum-Based Assessment Correlations.

** sig. at .01 level		LEAP 21 Total Math Score
Math Curriculum-based Measures	SSC	.228
	MSC	.415
	ACTC	.729**
	ACDC	.691**
Reading	MZR	.582**

Secondary analyses. Table 14 illustrates final grade correlations with standardized assessments and the curriculum-based assessments. Highest significant correlations were seen between LEAP 21 and final grade $r = .729$ $p<.01$, WJ III Broad score and final grades $r = .693$ $p<.01$ and of the math curriculum-based assessments, the ACTC = $r = .612$ $p<.01$.

Table 14. Fourth Grade Students' Final Grades, LEAP 21, Woodcock-Johnson III and Curriculum-Based Assessment Correlations.

** sig. at .01 level	LEAP 21 Total Math Score			
Students' Final Grades	.729**			
	Woodcock-Johnson III Math Measures			
	BS	Calc	FL	AP
Students' Final Grades	.693**	.628**	.229	.556**
	Math Curriculum-based Assessments			
	SSC	MSC	. ACTC	ACDC
Students' Final Grades	.198	.378	.612**	.595**

Table 15 shows correlations between teacher report data and the outcomes of the math curriculum-based assessments. The results indicate that the MSC assessment had the highest correlations with teacher report measures of applications and computation at r = .588 p<.01 and r = .578 p<.01 respectively.

Table 15. Teacher Report, Final Grade and Math Curriculum-Based Assessment Correlations for Fourth Grade.

** sig. at .01 level		Math Curriculum-based assessments				
		Final Grade	SSC	MSC	ACTC	ACDC
Teacher Report	Computation	.676**	.546**	.588**	.414	.452**
	Applications	.726**	.513**	.578**	.441**	.465**

Fifth Grade

Primary analyses. The correlation obtained between the ACTC and ACDC was r = .929 p<.01. Multiple skill computation assessment alternate forms correlations were: 1 and 2, r = .839 p<.01; 1 and 3, r = .753 p<.01 and 2 and 3, r = .819 p<.01. Maze alternate form correlations were: 1 and 2, r = .864 p<.01; 1 and 3, r = .819 p<.01 and 2 and 3, r = .812 p<.01.

Table 16 illustrates WJ III and math curriculum-based assessment correlations. Highest significant correlations occurred between the WJ III BS and the SSC, r = .800 p<.01 and ACTC, r = .790 p<.01.

Table 16. Fifth Grade Woodcock-Johnson III and Math Curriculum-Based Assessment Correlations.

** sig. at .01 level		Woodcock-Johnson III Math Measures			
		BS	Calc	FL	AP
Math Curriculum-based Measures	SSC	.800**	.494	.425	.704
	MSC	.684	.574	.700	.193
	ACTC	.780**	.446	.374	.586
	ACDC	.688	.395	.374	.125

Table 17 illustrates the correlations between the ITBS Test of Basic Skills and the fifth grade curriculum-based assessment scores. Highest significant correlations were observed between the ITBS TM and the ACTC, r = .688 p<.01.

Table 17. Fifth Grade ITBS Test of Basic Skills and Curriculum-Based Assessment Correlations.

** sig. at .01 level		ITBS Test of Basic Skills			
		IC	TM	MC	TR
Math Curriculum-based Measures	SSC	.548**	.620**	.447	.379
	MSC	.247	.340	.314	.189
	ACTC	.405	.688**	.531**	.244
	ACDC	.280	.578**	.508**	.179
Reading	MZR	.568**	.485	.432	.460

Secondary analyses. Table 18 illustrates the correlations between students' final grades, the standardized assessments and the math curriculum-based assessments. Results show that of the standardized assessments, the ITBS Total Math score had the highest significant correlation with students' final grade at r = .637 p<.01.

Table 18. Fifth Grade Students' Final Grades, ITBS Test of basic Skills, Woodcock-Johnson III and Curriculum-Based Assessment Correlations.

** sig. at .01 level	ITBS Test of Basic Skills			
	IC	TM	MC	TR
Students' Final Grades	.604**	.637**	.572**	.500**
	Woodcock-Johnson III Math Measures			
	BS	Calc	FL	AP
Students' Final Grades	.460	.183	.272	.464
	Math Curriculum-based Assessments			
	SSC	MSC	. ACTC	ACDC
Students' Final Grades	.739**	.500**	.645**	.528**

Of the math curriculum-based assessments, the SSC assessment had the highest correlation with students' final grades at r = .739 p<.01.

Table 19 illustrates correlations between teacher report data on applications and computation and the curriculum-based assessments. The results indicate that the single skill computation assessment had the highest significant correlations with teacher report data regarding student performance in computation, r = .635 p<.01 and applications, r = .659 p<.01.

Table 19. Teacher Report, Final Grade and Math Curriculum-Based Assessment Correlations for Fifth Grade.

** sig. at .01 level		Math Curriculum-based assessments				
		Final Grade	SSC	MSC	ACTC	ACDC
Teacher Report	Computation	.756**	.635**	.532**	.567**	.410**
	Applications	.770**	.659**	.416	.566**	.384

Face Validity: Math Curriculum-based Assessments

Table 20 illustrates that nearly all teachers preferred the newly created applications and computation math assessment over the SSC and MSC assessments as indicated on the Assessment Rating Profile 15. The top row denotes that two teachers from each grade (1^{st} through 5^{th}) evaluated the assessments. Row two illustrates teacher preference for the single skill computation assessment. Row three illustrates teacher preference for the multiple skill computation assessment and row four illustrates teacher preference for the applications and computation assessment (AC). Only one teacher, (first grade) did not prefer the AC assessment over the other math curriculum-based assessments. That teacher indicated that "the instructions were not stated for the computation questions", and "that would inhibit performance" for first grades students.

Table 20. Teacher Preference of Math Curriculum-Based Assessments.

Teachers	*1st grade*		*2nd grade*		*3rd grade*		*4th grade*		*5th grade*	
SSC	1.8	3.5	3.8	3.1	4.6	3.3	2.1	2	2.3	3.0
MSC	1.8	3.5	3.9	3.8	5.2	3.8	2.9	3.1	2.4	3.1
AC	6	3.3	4.8	6	6	4.3	5.9	6	5.8	4.1

Discussion

American schools have come under intense scrutiny over the last several years and have been asked to make substantial progress each year in the percentage of students who are proficient in reading and math, and in narrowing the achievement gap between advantaged and disadvantaged students. The primary accountability tool used to determine whether students are proficient is a yearly summative evaluation, or "state test" administered by states. From an instructional perspective, the state tests have little utility. That is, the tests are administered once late in the year and therefore provide no ongoing feedback to teachers about whether their instruction is producing desired student outcomes. Formative evaluation has traditionally been used for the purpose of ongoing evaluation of student learning. In mathematics instruction, for example, formative evaluation might evaluate progress on the many objectives and sub-objectives within the content standards proposed by states. Curriculum-based measurement is a well established tool for formative assessment. However, the problem with traditional CBM is that validity coefficients between CBM and state tests in mathematics has been moderate at best (Skiba, Magnusson, Martson, and Erickson, 1986; Martson, 1989; Putnam, 1989) making them possibly less accurate in predicting outcomes on state tests. The purpose of the present investigation was to develop and evaluate a set of math assessments designed to measure the type of application and problem-solving objectives required on state tests.

The "application" type assessment constructed for this study combined characteristics of CBM, CBA and CRA. The new assessment was evaluated in a manner such that the validity of the measure could be examined when used as a fluency measure or as an accuracy measure. The new assessments were derived from state standards and matched to the local district curriculum. Hence they allowed for an examination of student accuracy on large global outcome measures while also investigating accuracy (CBA) and performance on local curricula (CRA). An important question posed by this investigation concerns the benefit of taking the time to construct such assessments if similar correlations can be obtained through the assessment of traditional single skill or multiple skill computation assessments.

The discussion will consist of a summarizing the results, integration of the results into the existing literature, enumerating limitations of the study and finally proposing logical next steps for research on this topic. The data review will focus first on maze and MSC alternate form reliability followed by the ACTC and ACDC correlations. Next, the results pertaining to the curriculum-based assessment correlations with the WJ III, ITBS and LEAP will be presented. Finally, correlation data will be discussed relevant to the relationship between the math assessments and various teacher report measures including students' final grades, teacher report of student performance, and teacher preference for math assessments.

Maze, Multiple Skill Computation, and Applications and Computation Correlations

Moderate to high alternate form correlations were obtained for the maze reading passages across the grades ranging from .583 (1^{st} grade) to .925 (2^{nd} grade) with median correlations ranging from .720 (for third grade) to .900 (for second grade) suggesting that the passages used were comparable regarding grade-level material. Moderate to high alternate form correlations were obtained for the MSC assessments, ranging from .681 (3^{rd} grade) to .859 (2^{nd} grade) with median correlations ranging from .751 (for third grade) and .834 (for second grade).

The applications and computation assessment scores yielded very high correlations between the ACTC and ACDC, ranging from .869 to .987. This suggests that the relationship between the number of total problems successfully completed was highly correlated with the total number of digits correct. This relationship is not surprising, but it does suggest that students who attempted to complete a problem finished the problem correctly more often than not. The high correlation between the two measures might allow for a quick, easy scoring procedure that possibly reduces errors associated with counting digits correct.

Curriculum-Based Assessments and Criterion Assessment Correlations

Woodcock-Johnson III. The analysis of AC assessment scores yielded significant correlations the WJ III BS in four of the five grades (grades one through four), though similar, significant correlations were obtained by the MSC in three grades (grades one through three). The AC scores also yielded statistically significant correlation coefficients with the AP sub-tests of the WJ III for four of the five grades (grades two, three, four and five). As anticipated, in comparison with AC scores, the MSC and SSC assessments were more likely to have statistically significant correlations with the WJ III FL measure. Only first grade AC scores were significantly correlated with the FL measure.

Iowa Test of Basic Skills. Significant correlations were obtained between ITBS math measure TM for the AC assessment in three of the three grades assessed with the ITBS (grades two, three and five), whereas of the computation assessments (i.e., SSC and MSC) only the SSC assessment scores yielded a significant correlation with TM and only in one grade (fifth). Similarly, the AC assessment yielded significant correlations with the MC score of the ITBS (all three grades), whereas the SSC and MSC assessments only yielded significant correlations in two of the three grades, both were the MSC assessment, not the SSC.

Louisiana Education Program for the 21st Century. Of the math CBA assessments, only the AC assessment yielded significant correlations with the LEAP total math score. Interestingly, the MZR assessment score also produced significant correlations with the LEAP total math score.

Final Math Scores and Teacher Report Data. Of the math assessments, only the AC assessment scores yielded significant correlations with students' final grades in all five grades. The SSC assessment score yielded significant correlations in four of five and the MSC score yielded significant correlations in three of five grades.

The WJ III sub-tests yielded only six significant correlations of the 20 possible combinations (i.e., BS, CALC, FL, and AP in first grade, second grade, third grade, fourth grade and fifth grade). They were BS in third and fourth grades, CALC in fourth grade, FL in first grade and AP in third and fourth grades. No significant correlations were observed in second or fifth grades.

The ITBS TM and MC scores yielded significant correlations in all three grades and the LEAP TM score yielded a significant correlation with students' final grades in fourth grade. Only the AC assessment scores yielded significant correlations with both teacher report of student performance on computation and applications in all five grades. The SSC and MSC assessment scores did yield significant correlations in four of the five grades.

Finally, teachers overwhelmingly favored the AC assessment. That is, the mean scores on the ARP for the SSC, MSC and AC assessments were 2.95, 3.35 and 5.22 (range 1 – 6, where 1 represented a less favorable score and 6 represented a highly favorable score), respectively. Additionally, four of ten teachers agreed with all 15 statements on the ARP.

Overall, the results suggest that the newly constructed math application assessments were more likely to yield significant correlations with the criterion assessments than the traditional M-CBM assessments, (i.e., SSC and MSC). Also, the application assessments were preferred by teachers over the SSC and MSC assessments. Reading is an important component of math in that the maze reading passages were more likely to yield significant correlations with the ITBS measures than the SSC and MSC. Maze reading also yielded a statistically significant association with the LEAP TM score, whereas neither the SSC nor the MSC yielded a significant correlation.

Relevance of Results for the Extant Literature

Allinder and Oats, (1997) suggested that teachers are more likely to use CBM assessments they like than ones they do not like. This is an important point when considering that progress monitoring of basic skills increases student performance (i.e., performance indicator) and assists teachers in intervention planning (i.e., skills analysis). As such, if preference predicts use, then teachers may be more likely to monitor student performance on applications problems if they were provided with an assessment such as the one used here. In other words, teachers would be more likely to use this CBA assessment than the SSC assessment or the MSC assessments, when investigating how their students will perform on state tests.

Next, the results obtained in the present investigation reveal that the median score obtained on the MSC assessment produced similar results as the one SSC assessment. As suggested by Hintze, Christ and Keller (2002), the median score of three alternate forms should be used when comparing MSC assessments to the criterion-assessments. The median score is suggested because algorithms will change on MSC assessments due to the high degree of problem variability when assessing the various skills (addition, subtraction, multiplication or division). As such, one assessment may reveal that a student made few errors, and therefore requires little intervention, and another assessment might produce many errors, suggesting the need for significant intervention. Though the MSC alternate form reliability was significant within grade, there was also considerable within grade variability between the scores on the criterion assessments and the three alternate forms. Taking the median score for each grade resulted in correlation coefficients between the MSC and the criterion assessments similar to obtained with the SSC assessment. These results replicate the findings by Hintze, Christ and Keller.

The results of the present study also support Thurber, Shinn and Smolkowski (2002) regarding the reason correlation coefficients between math CBM and the criterion assessments have historically been low. The authors suggest that the extant literature has reported low correlation coefficients because the curriculum-based assessments had: 1) limited content validity, and 2) and the criterion assessments where heavily laden with reading (e.g., instructions and word problems). The present data support those suggestions. For example, when addressing the issue related to content validity, previous studies assumed that computation and applications were highly related constructs and therefore the content on the CBM assessments (i.e., computation problems) reflected the content of the criterion assessments. This was a faulty assumption as demonstrated by Thurber, Shinn and Smolkowski (2002). Their results revealed that applications and computation problems were distinct constructs, though their results do reveal low correlations between the two constructs, and are therefore related. The primary investigators in the present study did not assume that performance on computation problems would be highly correlated with performance on the criterion assessments; rather they correctly hypothesized that performance on computation problems would be correlated with performance on computation sub-tests of the criterion assessments. Additionally, the present study demonstrated that performance on an assessment designed to measure applications problems would be significantly correlated with the criterion assessments that also measure applications. As such it should not be surprising that the AC assessment yielded significant correlations with the LEAP and ITBS. For example, the ITBS and LEAP assessments include items from content standards in geometry, measurement, number systems, data analysis and patterns. Therefore, the AC assessment scores should be correlated with the criterion assessments used in this investigation. Additionally, given that the LEAP and ITBS contain directions and word problems, it should not be surprising that the MZR assessment was better at predicting performance than the SSC and MSC assessments.

As expected, the SSC and MSC assessments were better at predicting performance on the WJ III FL measure than the AC assessment and MZR assessments. This again confirms the conclusions offered by Thurber, Shinn and Smolkowski (2002) regarding computation and applications problems as two distinct constructs. Furthermore, many of the previous investigations thus far have compared fluency based computation assessment scores to scores obtained on assessments such as the Metropolitan Achievement Test (MAT) and yielded low correlation coefficients. As such, those studies have attempted to demonstrate a high correlation between two the distinct constructs offered by Thurber, Shinn and Smolkowski (2002).

To date, the only investigation reviewed in the extant literature that included a broad range of applications problems (e.g., measurement, geometry, algebra, patterns) and produced high correlation coefficients with a the criterion assessment, the Comprehensive Test of Basic Skills (CTBS), was that conducted by Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek and Stecker (1994). The authors demonstrated that when a math curriculum-based assessment contained material consistent with state content standards prescribed by the local curricula, then respectable validity coefficients can be obtained (range from .64 to .81) in grades two, three and four. However, the authors did not report if those correlation coefficients indicated a stronger degree of association than those yielded on SSC or MSC assessments. Rather the authors compared the slope (rate of acquisition) between applications problems and computation problems across the year in second, third and fourth grades. The results showed that the rate of growth in second grade on the applications problems was .40, where as the rate of growth on the computation problems was .25. The authors concluded that the rate of growth on the applications problems was not necessarily related to growth in math skills as much as it was related to growth in reading, which was required to answer the math applications problems. In other words, second grade students had a higher rate of growth on the math applications assessment in second grade, because in second grade their reading became more proficient. The data in the present investigation support the importance of reading given that the maze reading passages had higher correlations with the criterion-measures than the SSC or MSC assessments, therefore confirming that reading is a prerequisite skill needed on assessments containing applications problems. Interestingly, in the Fuchs et al, investigation, students' computation and applications slopes were very similar in fourth grade, .70 and .69, respectively. This suggests as students become more proficient in reading, that growth on applications problems will be similar to growth on computation problems. Lacking from the Fuchs et al. investigation was data demonstrating that growth (i.e., slope) in computation problems would be predictive of student growth in applications problems. The data from Thurber, Shinn and Smolkowski (2002) suggests that this relationship may be weak.

Further investigation in the present study reveals that the relationship between performance on the AC assessment and the SSC and MSC is low (Person product correlations between .127 and .378). These results might suggest, as Thurber, Shinn and Smolkowski demonstrated, that the constructs of applications and computation are distinct, and therefore to investigate how well a student will perform on a the criterion assessment, applications problems (e.g., algebra, data analysis, geometry, measurement) may need to be included on the assessment.

Limitations

Though the findings in the present investigation are promising, there are several limitations. First, given the advantages of fluency for progress monitoring, it was disappointing that the present investigation did not produce fluency scores that yielded significant correlations with the criterion assessments. One potential reason for this result was the amount of time students were allotted for the fluency measure (five minutes). Fuchs et al allocated up to eight minutes to complete six fewer problems on the grade-level assessments. Given that the AC assessment contained reading directions and word problems, future research may want to incorporate increases in the amount of time to determine if eight or ten minutes could yield better results. However, no student required more than 30 minutes to complete the accuracy component of the assessment, and most were finished in less than 20 minutes. That is important when considering the findings. If an assessment such as this could predict end of year performance, then 20 minutes two or three times a year might be a worthwhile endeavor.

Next, this assessment was given late in the year, after the criterion assessments were administered. Hence it is not known the extent to which the present assessment might be useful from a formative evaluation perspective. Future research may need to investigate two issues pertaining to this limitation. First, is the applications assessment sensitive to instruction? Research by Fuchs et al, suggests that it is. Second, would performance at the beginning or the middle of the year produce the same significant correlations with the criterion assessments and final grades such that modifications to instruction could reveal students in need of instructional modifications? Again, the results of Fuchs, et al suggest that would be possible, though it is not demonstrated by this investigation. Future research would administer this assessment at the beginning, middle and end of the year to chart student growth and indicate where instructional modifications are needed.

In order to proceed with the recommendation above, alternate forms may be beneficial. Therefore, future research would replicate the procedure described here such that alternate forms could be generated. Though the process to generate the first assessment took more than an hour and the validation process longer still, the overall time decreased with each new assessment because the materials are already assembled and practice with the process decreases the time needed to construct an assessment. Replication of this procedure would not require the same time given that the procedures are already identified.

Procedural replications are also recommended to investigate the outcomes in different regions, with different curricula, with different populations and with different criterion assessments. If performance on the criterion assessments is not predictable by the curriculum-based assessment procedures currently employed, then either this procedure, the one described by Fuchs et al, or another procedure may be warranted in order to assess the generalizability of applications assessments with the various populations, curricula and the criterion assessments. Replication of this procedure would be desirable in order to determine whether the results obtained here, would be similar given different professionals constructing the assessments with the various math curricula used throughout the country. If so, then future research regarding the assessment of application problems may be forthcoming, thereby assisting educators to prepare students for state tests and success in math problem-solving skills.

Finally, the sample size did not permit the use of multivariate procedures which might have allowed for a deeper understanding of the relations among variables. Future research might administer these assessments to larger schools or multiple schools therefore enabling multivariate analysis.

Although this study did have limitations, the results suggest that the methodology for constructing the applications grade-level assessments is promising. Additionally, the validity coefficients between the applications assessments and the criterion assessments were much higher than previous math curriculum-based assessments studies that have only investigated the relationship between computation and state tests. Finally, the teachers surveyed were encouraged by an assessment which was constructed from their curriculum, administered in less than 30 minutes and yielded moderate to high correlation coefficients. Though this does not suggest that the teachers will use this assessment to monitor progress or inform instruction, it is a step in the right direction.

References

Allinder, R. M. (1996). When Some is not Better than None: Effects of Differential Implementation of Curriculum-Based Measurement. *Exceptional Children, 62*, 525-535.

Allinder, R. M. and Beckbest, M. A. (1995). Differential Effects of Two Approaches To Supporting Teachers' Use of Curriculum-Based Measurement. *School Psychology Review, 24*, 287-298.

Allinder, R. M. and Oats, R. G. (1997). Effects of Acceptability on Teachers' Implementation of Curriculum-Based Measurement and Student Achievement in Mathematics Computation. *Remedial and Special Education, 18*, 113-120.

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., Slider, N. J. and Williams, K. L. (2004). Examining the Incremental Benefits of Administering a Maze and Three versus One Curriculum-Based Measurement Reading Probes when Conducting Universal Screening. *School Psychology Review, 33*, 218-233.

Blankenship, C. (1985). Using Curriculum-Based Assessment Data to Make Instructional Decisions. *Exceptional Children, 52*, 233-238.

Blankenship, C. and Lilly, S. (1981). Mainstreaming Students with Learning and Behavior Problems. *Techniques for the Classroom Teacher*. New York: Holt, Reinhart and Winston.

Deno, S. L. (1987). Curriculum-Based Measurement. *Teaching Exceptional Children, 20,* 41.

Deno, S. L. and Mirkin, P. (1977). *Data-Based Program Modification: A Manual*. Reston, VA. Council for Exceptional Children.

Espin, T. A. and Tindal, G. (1998). Curriculum-Based Measurement for Secondary Students. In M. R. Shinn, Mark (Ed.), *Advanced Applications of Curriculum-Based Measurement*. New York, NY: Guilford Press.

Foegen, A. and Deno, S. L. (2001). Identifying Growth Indicators for Low-Achieving Students in Middle School Mathematics. *The Journal of Special Education, 35,* 4-16.

Frisby, C. (1987). Alternative Assessment Committee Report: Curriculum-Based Assessment. *CASP Today, 36*, 15-26.

Fuchs, L. S., Fuchs, D. and Hamlett, C. L. (1989c). Effects of Instrumental Use of Curriculum-Based Measurement to Enhance Instructional Programs. *Remedial and Special Education, 10*, 51-59.

Fuchs, L. S., Fuchs, D., Hamlett, C. L. and Appleton, A. C. (2002). Explicitly Teaching for Transfer: Effects on the Mathematical Problem-Solving Performance of Students with Mathematics Disabilities. *Learning Disabilities Research and Practice, 17*, 90-106.

Fuchs, L. S., Fuchs, D., Hamlet, C. L. and 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*, 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubek, P. and Stecker, P. M. (1994). Technical Features of a Mathematics Concepts and Applications Curriculum-Based Measurement System. *Diagnostique, 19*, 23-49.

Gickling, E. and Havertape, J. (1981). Curriculum-Based Assessment (CBA). Minneapolis: National School Psychology Inservice Training Network.

Gickling, E. and Thompson, V. (1985). A Personal View of Curriculum-Based Assessment. *Exceptional Children, 52*, 205-218.

Good, R. H. and Kaminski, R. A. (1996). Assessment for Instructional Decisions: Toward a Proactive/Prevention Model of Decision Making for Early Literacy Skills. *School Psychology Quarterly, 11*, 1-11.

Hagris, C. H. (1987). *Curriculum-Based Assessment: A Primer*. Springfiled, Il: Charles C. Thomas

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*, 514-528.

Hintze, J. M., Owen, S. V., Shapiro, E. S. & Daly, E. J. (2000). Generalizability of Oral Reading Fluency Measures: Application of G Theory to Curriculum-Based Measurement. *School Psychology Quarterly, 15*, 52-68.

Howell, K. W., Fox, S. L. and Morehead, M. K. (1993). *Curriculum-Based Evaluation: Teaching and Direction Making*. Pacific Grove, CA: Brooks/Cole.

Howell, K. W., Zucker, S. H. and Morehead, M. K. (2000). Multilevel Academic Skills Inventory. Western Washington University, Bellingham, WA: Applied Research and Development Center.

Idol-Maestas, L. (1983). *Special Educators Consultation Handbook.* Rockville, Md: Aspen

Marston, D. (1989). A Curriculum-Based Measurement Approach to Assessing Academic Performance: What is it and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.

Martson, D., Deno, S. L. and Kim, D. (1995). Comparison of Reading Intervention Approaches for Students with Mild Disabilities. *Exceptional Children, 62*, 20-37.

Math Central. (2003) Houghton Mifflin. Geneva, IL.

Micro Power & Light Co. (2000). Readability Calculations. [Computer software]. Retrieved from http://www.micropowerandlight.com/rd.html

National Center for Education Statistics (n.d.). *The Nation's Report Card,* Retrieved from http://nces.ed.gov/nationsreportcard/

National Center for Education Statistics (n.d.). *Third International Mathematics and Science Study*. Retrieved from http://nces.ed.gov/timss/

Putnam, D. (1989). The Criterion-Related Validity of CBM Measures of Math. Unpublished Master's Thesis, University of Oregon, Eugene.

Salvia, J. and Ysseldyke, J. E. (1991). *Assessment* (5th ed). Boston: Houghton Mifflin.

Schoolhouse Technologies Inc. (2003). Mathematics Worksheet Factory Deluxe (Version 3.0) [Computer software]. Retrieved from http://www.shcoolhousetechnologies.com

Shapiro, E. S. (1996). *Academic Skills Problems: Direct Assessment and Intervention.* (2nd ed.) New York: Guilford.

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

New York: Gilford.

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

Shinn, M.R., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*(2), 31-38.

Shinn, M.R., Good, R.H., Knutson, N., Tilly, W.D., & Collins, V.L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

Silbert, J., Carnine, D. and Stein, M. (1990). Direct Instruction Mathematics (2nd ed.). Columbus, OH. Merrill.

Skiba, R., Magnusson, D., Martson, D. and Erickson, K. (1986). *The Assessment of Mathematics Performance in Special Education: Achievement Tests, Proficiency Tests, or Formative Evaluation?* Minneapolis: Special Services, Minneapolis Public Schools.

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*(4), 498-513.

VanderHeyden, A. M, Witt, J. C. and Naquin, G. (2001). The Reliability and Validity of Curriculum-Based Measurement Readiness Probes for Kindergarten Students. *School Psychology Review, 30,* 363-382.

Witt, J. C. and Elliot, S. N. (1985). Acceptability of Classroom Intervention Strategies. In T. R. Kratochwill (Ed.), *Advances in School Psychology.* Hillsdale, NJ: Erlbaum.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20*, 510-517.

Appendix A

Appendix A. General Curriculum Standards Reported by Seven Randomly Chosen States and Their Definitions.

States	Number Sense	Data Analysis	Patterns	Algebra	Geometry	Measurement
Louisiana	x	x	x	x	x	x
Arizona	x	x	x	x	x	x
Massachusetts	x	x	x	x	x	x
South Carolina	x	x		x	x	x
Alaska	x		x	x	x	x
Connecticut	x	x	x	x	x	x
Mississippi	x	x	x	x	x	x

Definitions of Six Standards

Number Sense. The understanding of numbers, their relation to each other, and their real world usage. Includes awareness of measuring and counting, whole numbers, fractions and decimals and the understanding of number size.

Data Analysis. Includes data collection, probability, statistics and graphing.

Patterns. Use pattern recognition to solve math problems.

Algebra. Identify inverse relationships like those found in fact families. Understand =, <, & > symbols.

Measurement. Compare volume, weight, length of two or more items. Measure in inches or metrics. State time of day as seen on an analog clock. State days of week and moths of year in order. Count money.

Geometry. Recognize and classify shapes and objects. Use words that indicate position (such as: next to, beside, between & across). Draw lines through shapes to indicate parts.

Appendix B. Math Applications and Computation Assessments.

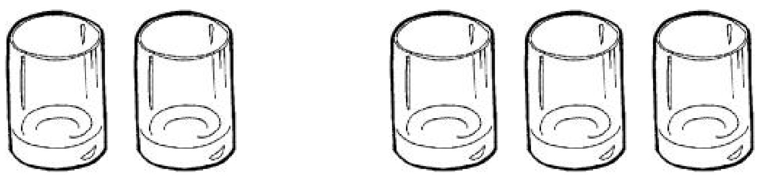
First Grade
Math Assessment

Teacher________________________
School_________________________
Student________________________
Date___________________________

N1. 76 – 10 =

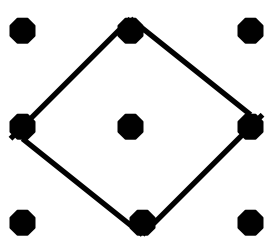
a) 56
b) 86
c) 66

A1. Write the number sentence.

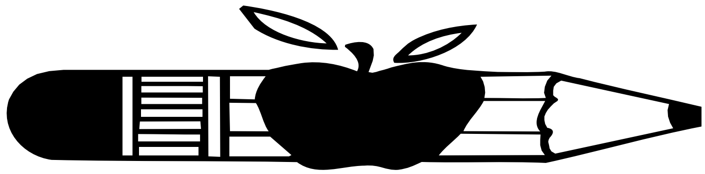


________________+____________________=__________

G1. Use the figure. How many corners does the figure have?____________

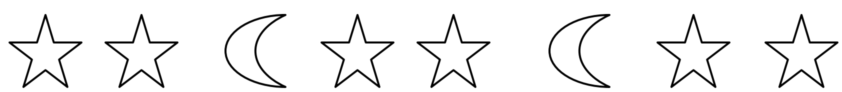


M1. Measure. Use an inch ruler.

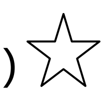


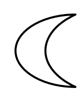
About how many inches?_________

P1. Look at the picture. Answer the question.

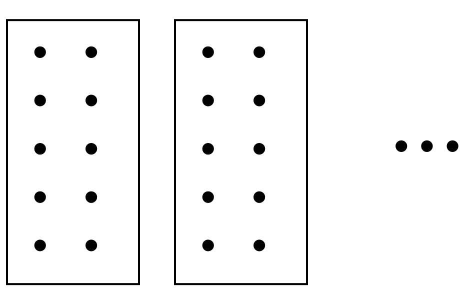


What comes next?

a) 

b) 

D1. Write how many tens and ones:

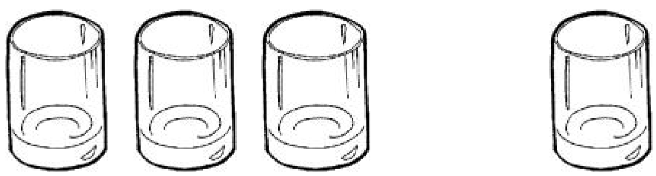


Tens	Ones

N2. 38 + 22 =

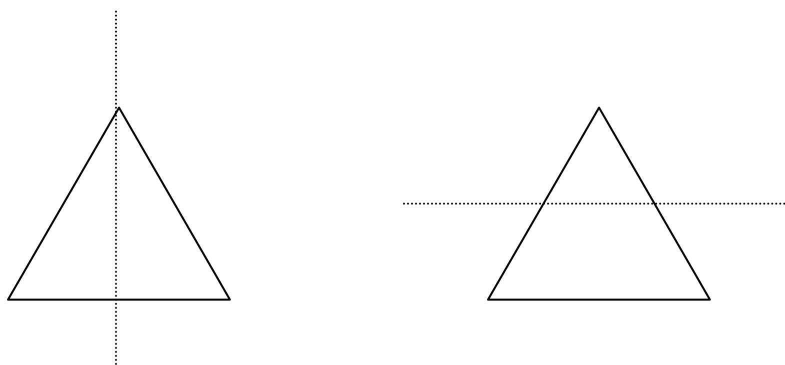
a) 56
b) 80
c) 60

A2. Write the number sentence.

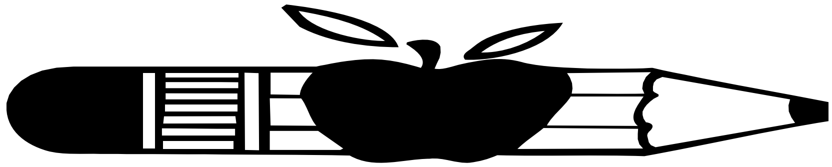


______________________+__________=__________

G2. Circle the figure that will fold in half.



M2. Measure. Use an inch ruler.

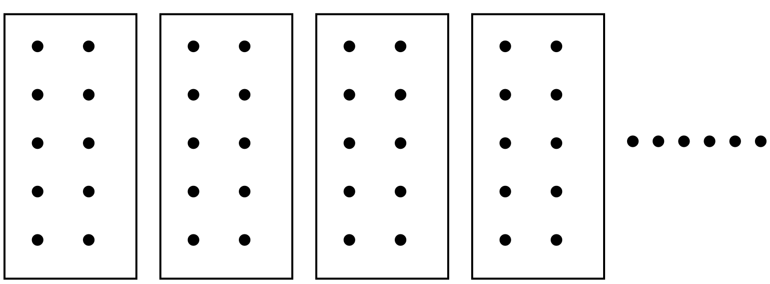


About how many inches?_________

P2. What number is between 7 and 9?

a) 6
b) 8
c) 10

D2. Write how many tens and ones:



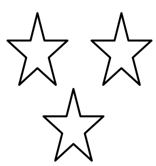
Tens	Ones

N3.

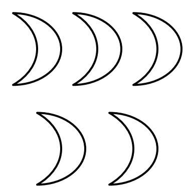
$$\begin{array}{r} 66 \\ +34 \\ \hline \end{array}$$

a) 90
b) 100
c) 92

A3. Find the set that has more. Circle the correct answer.

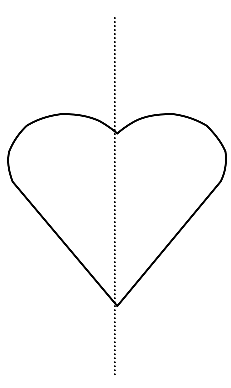


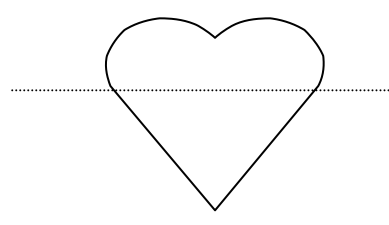
Set A



Set B

G3. Circle the figure that will fold in half.



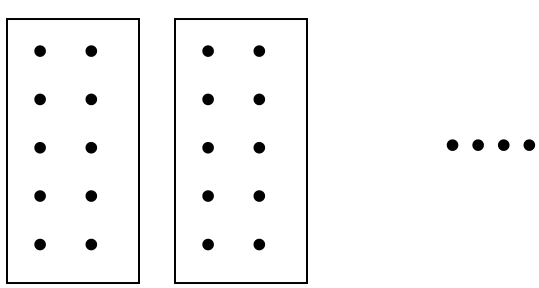


M3. Circle the heavier object.

P3. What number comes just before 6?

a) 4
b) 5
c) 7

D3. Write how many tens and ones:



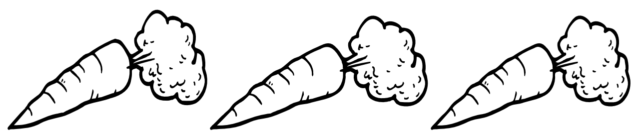
a) 14
b) 30
c) 24

N4. There are 9 apples  on the desk. Kim takes away five of them. How many apples  are left on the desk?

a) 9 – 4 = 5
b) 9 – 5 = 4
c) 5 – 4 = 9

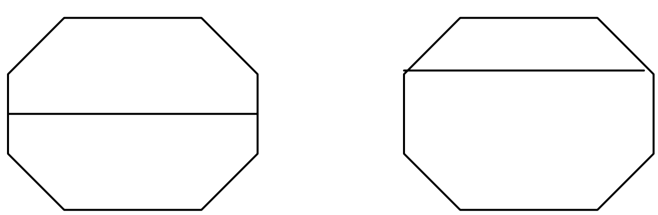
A4. Find the set that has fewer.

a)



b)

G4. Circle the figure that shows equal parts.



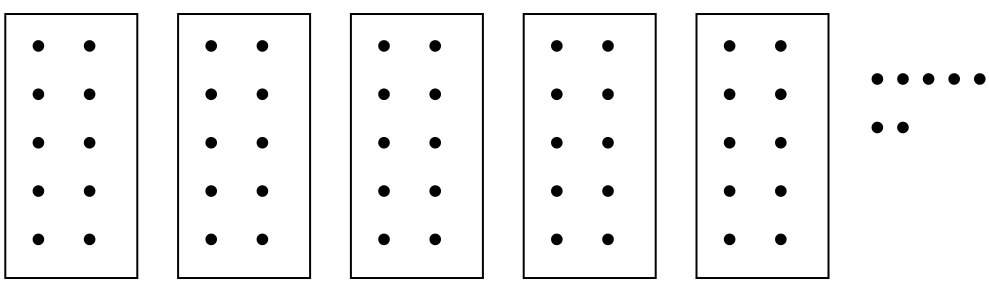
M4. Circle the temperature.

P4. Which number comes just after 3?

a) 4
b) 1
c) 2

D4. Choose how many.

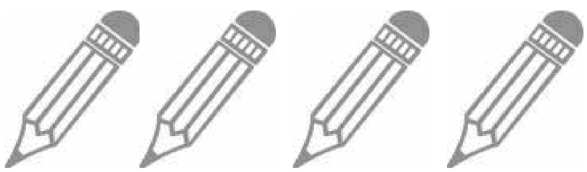


a) 47
b) 57
c) 60

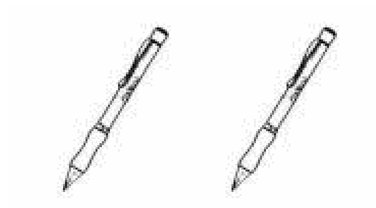
N5. There are 19 balls in a bag. Bob takes out 10 of them. How many balls are in the bag now?

a) 8 balls
b) 9 balls
c) 10 balls

A5. Find the set that has more.

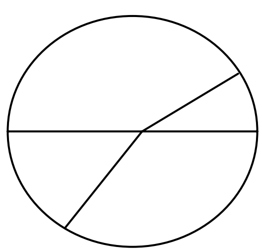


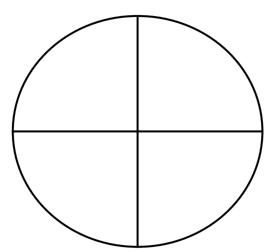
a)



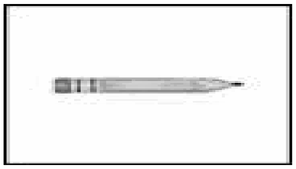
b)

G5. Circle the figure that shows equal parts.





M5. Circle how you measure
- how long is the pencil?



a)

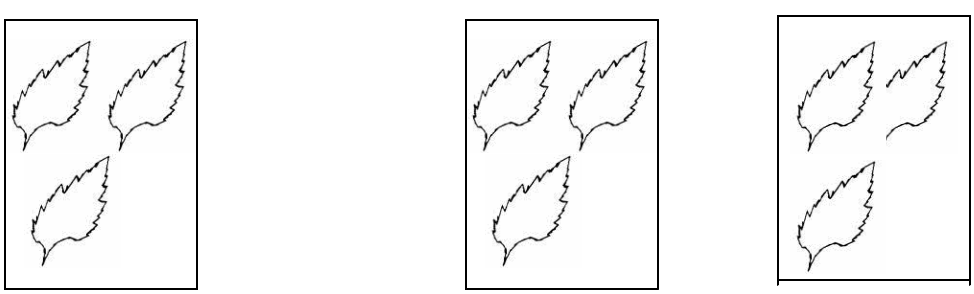
b)

P5. Find the missing number.

_____, 56, 57

a) 55
b) 54
c) 57

D5. Write how many.
How many are in each set? ___________



N6. Jill and Lisa have the same number of dolls. Lisa gives away 4 dolls. Who has more dolls?

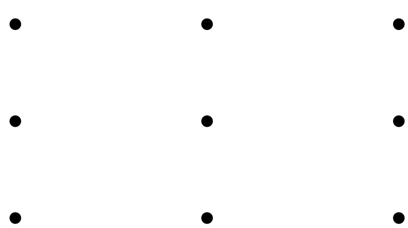
a) Jill
b) Lisa

A6. Find the set that has fewer.

a)

b)

G6. Draw a triangle using the dots.



M6. Circle how you measure
-How much does the milk container hold?

a)

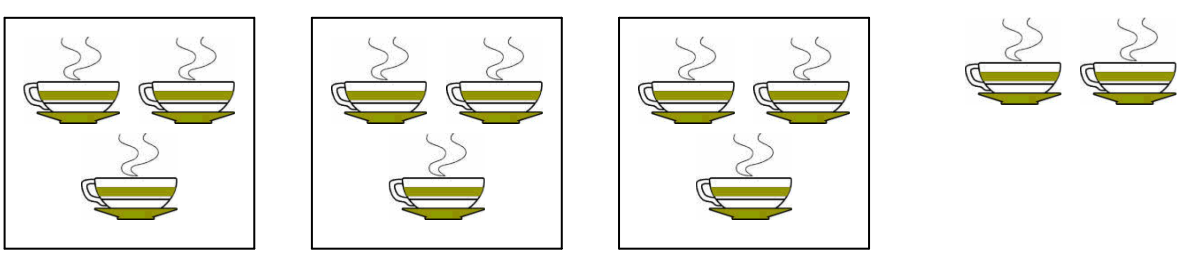
b)

P6. Find the missing number.

71, _____, 73

a) 70
b) 74
c) 72

D6. Write how many.
-how many are left over? _____



Second Grade
Math Assessment

Teacher_______________________
School________________________
Student_______________________
Date__________________________

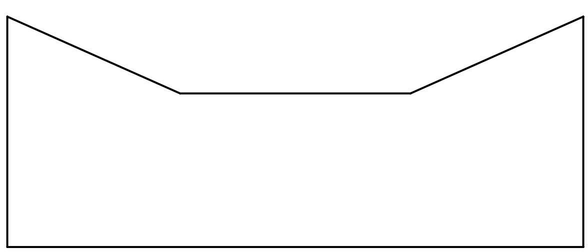
N1. 11 – 5 = _____

A1. Pick > , < , or =

45 $\square$ 54

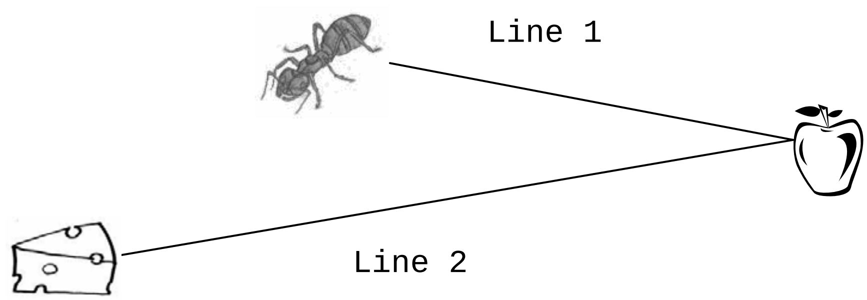
a) >
b) <
c) =

G1. Write how many sides.



_____ sides

M1. Use an inch ruler to measure.
Write the answer.

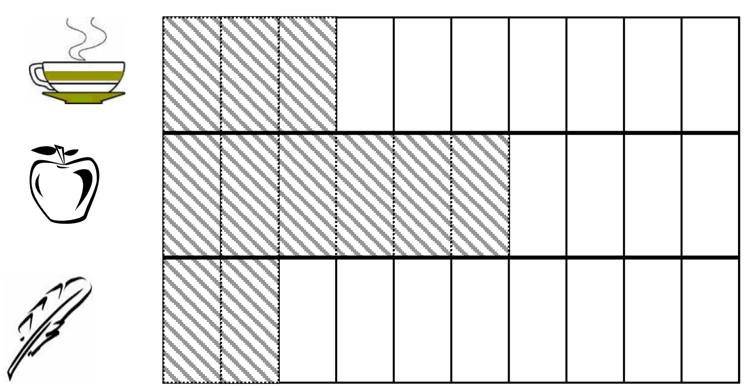


1. Line 1 is about _____ inches.
2. Line 2 is about _____ inches.

P1. Write the missing number:

3, _____, 5, 6

D1. Use the graph.



How many more apples than feathers? _____

N2.

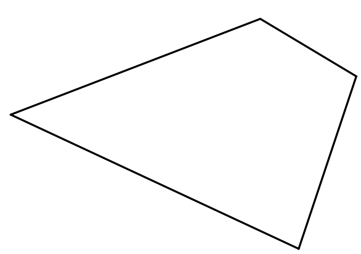
$$\begin{array}{r} 8 \\ 3 \\ +\quad 1 \\ \hline \end{array}$$

A2. Pick > , < , or =

74 $\boxed{}$ 77

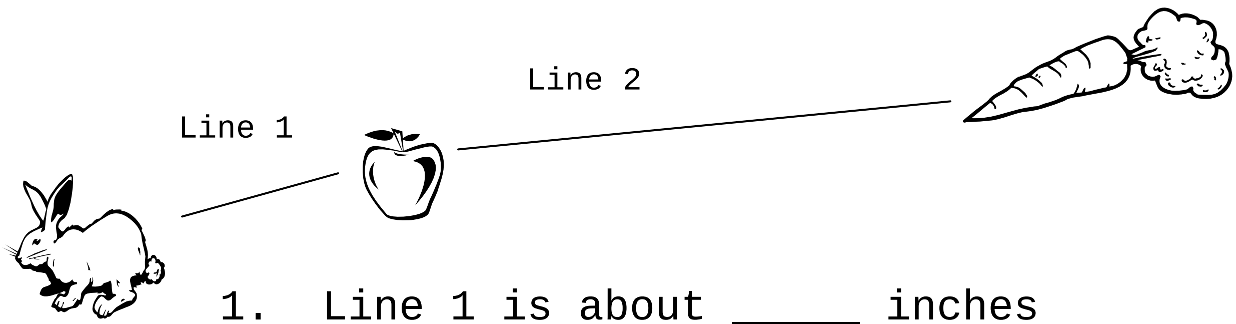
a) <
b) >
c) =

G2. Write how many corners.



______ corners

M2. Use an inch ruler to measure.
Write the answer.



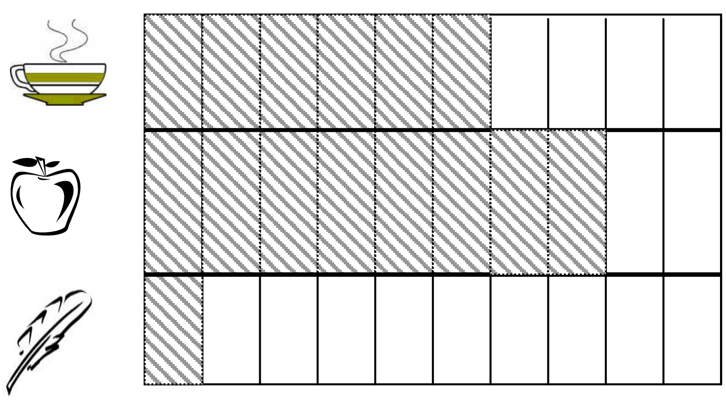
1. Line 1 is about _____ inches

2. Line 2 is about _____ inches

P2. What the missing number.

8, _____, 6

D2. Use the graph.



How many fewer feathers than apples? _____

N3. Write the number that completes the number sentence.

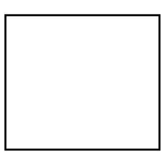
$$4 + \_\_\_\_\_ = 11$$

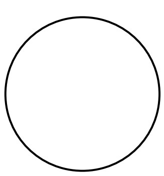
A3. Pick > , < , or =

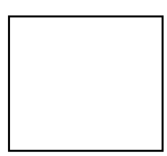
56 $\square$ 52

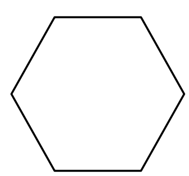
a) <
b) >
c) =

G3. Circle the one that is the same.



a) 

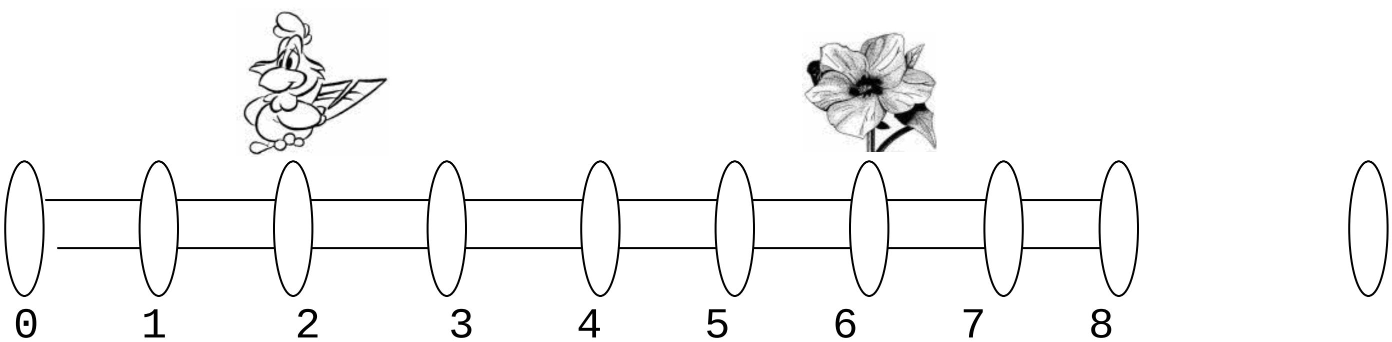
b) 

c) 

M3. Circle the best estimate.

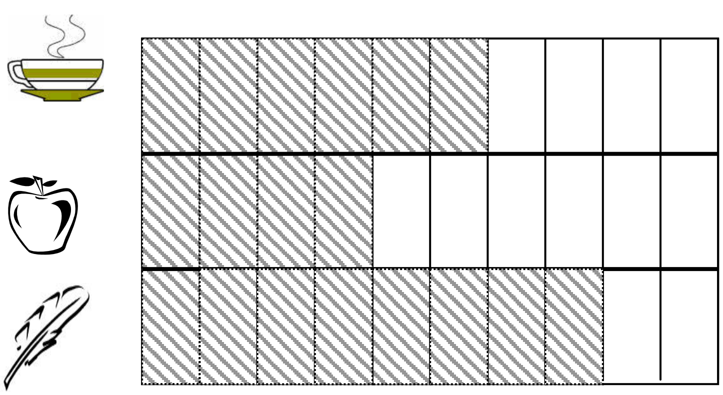
a) more than a pound
b) about a pound
c) less than a pound

P3. What numbers are between the bird and the flower?



_____, _____, _____.

D3. Use the graph.

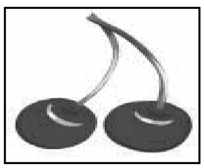


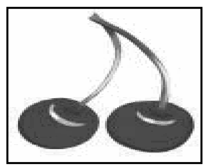
How many more cups than apples? _____

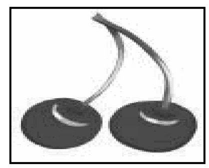
N4. On the field day 100 children had pickles. During the day 63 children ate their pickle. How many children still have pickles?

_____ children

A4. Write how many.



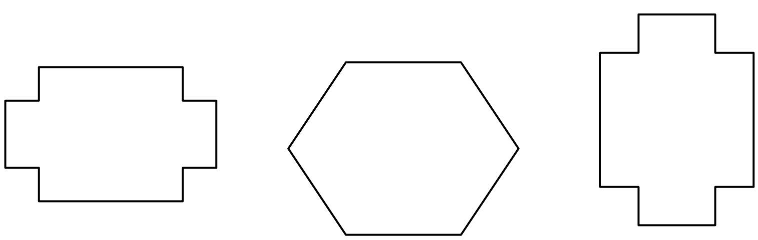




________ + ________ + ________ = ________

________ sets of ________ = ________

G4. Circle the one that is different.



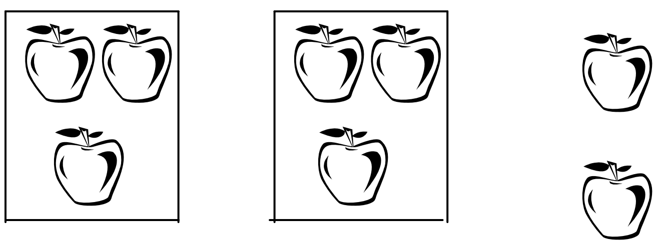
M4. Circle the best estimate.

a) more than a gallon
b) about a gallon
c) less than a gallon

P4. Write the missing number.

4, 8, 12, ______, 20.

D4. Find how many.



-How many apples?

a)6

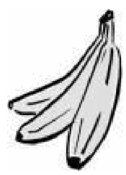
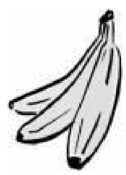
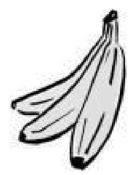
b)8

c)10

N5. There are 5 plates on the table. On each plate are 3 grapes. How many grapes are there?

a) 15 grapes
b) 8 grapes
c) 2 grapes

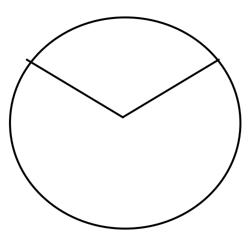
A5. Write how many.

________ +________+ ________ = ________

________sets of ________ = ________

G5. Color the figure to show the fraction.
- two thirds 2/3.



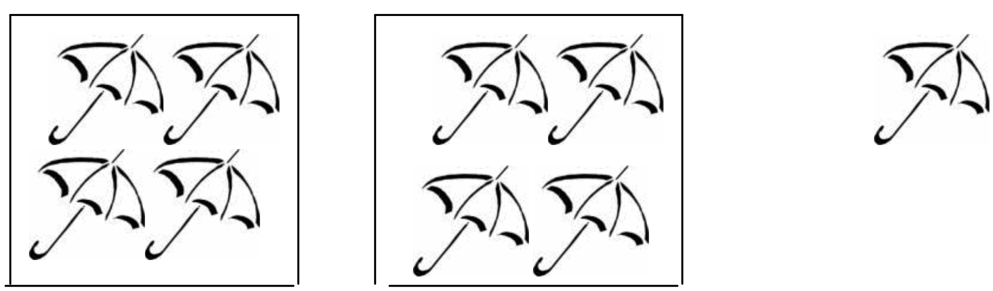
M5. Circle the best estimate.

a) more than a liter
b) about a liter
c) less than a liter

P5. Write the missing number.

30, _____, 45, 50

D5. Find how many.



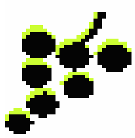
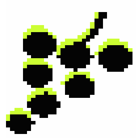
-How many sets?

a) 1
b) 2
c) 3

N6. Tanisha has $1.50 to spend at the movies. April has $2.00 to spend. How much do Tanisha and April have to spend altogether?

a) $2.50
b) $3.50
c) $4.50

A6. Write how many.

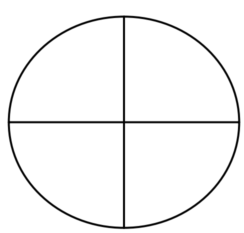
 

________+ ________ = ________

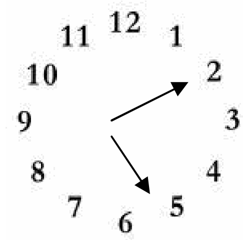
________ sets of ________ = ________

G6. Color the figure to show the fraction.

- one fourth 1/4



M6. Find the time.

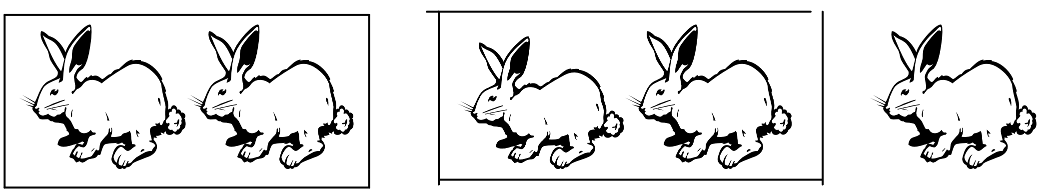


a) 5:50
b) 5:10
c) 6:10

P6. Write the missing number.

_____, 50, 60, 70.

D6. Find how many.



-How many left over?
a) 0
b) 1

Third Grade
Math Assessment

Teacher___________________________
School____________________________
Student___________________________
Date______________________________

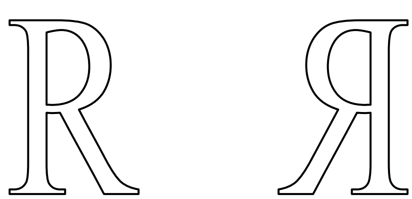
N1. Write the number in standard form.

500 + 60 + 3

A1. Complete the number sentence.

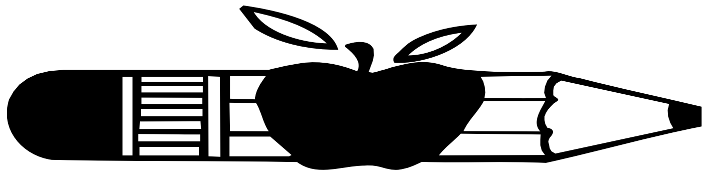
5 + ________ = 12

G1. Choose slide, flip, or half-turn for the set of figures.



a) slide
b) flip
c) half-turn

M1. Measure. Use a ruler to measure to the nearest inch.



Nearest half inch? _______

P1. Write the product that the symbol stands for.

1	*2*	*3*	*4*
2	*	*6*	*8*
3	*6*	*9*	*12*
4	*8*	*12*	*16*

* = _______

D1. Use the chart.

How many more students picked dogs rather than cats as their favorite pets? _______

Favorite Class Pets.

<table>
<tr><td>Pet</td><td>Tally</td></tr>
<tr><td>Dog</td><td>~~||||~~ |</td></tr>
<tr><td>Cat</td><td>|||</td></tr>
<tr><td>Fish</td><td>||</td></tr>
</table>

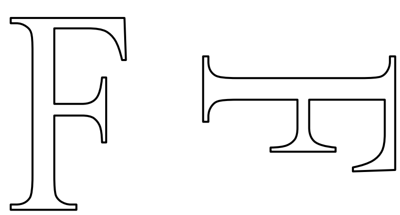
N2. Find the sum.

$$\begin{array}{r} 436 \\ +\ 77 \\ \hline \end{array}$$

A2. Complete the number sentence.

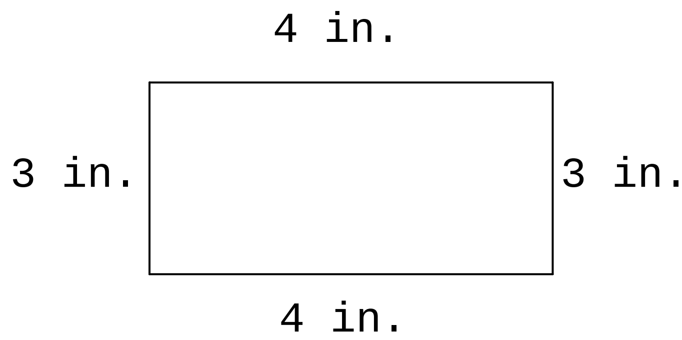
_______ + 8 = 17

G2. Choose slide, flip, or half-turn for the set of figures.



a) slide
b) flip
c) half-turn

M2. Write the perimeter.



____

P2. Write the product that the symbol stands for.

1	*2*		*4*
2	*4*		*
		9	*12*
4		*12*	

* = ________

D2. Use the chart.

How many students took part in the survey? ______

My Favorite Color

<table>
<tr><th>Color</th><th>Tally</th></tr>
<tr><td>Red</td><td><s>||||</s> |</td></tr>
<tr><td>Blue</td><td>||</td></tr>
<tr><td>Black</td><td><s>||||</s> |||</td></tr>
</table>

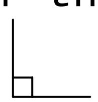
N3. Find the difference.

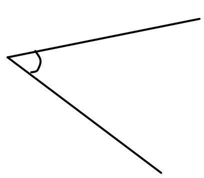
$$\begin{array}{r} 335 \\ -\ 56 \\ \hline \end{array}$$

A3. Write the number sentence that is missing from the fact family.

7 + 2 = 9
2 + 7 = 9
9 - 7 = 2

____

G3. Choose if the angle is <, >, or =
a right angle 



a) <
b) >
c) =

M3. Write the correct unit to measure the following.

-water in a bathtub

a) ounce
b) cup
c) gallon

P3. Write the product that the symbol stands for.

1			*4*
	4	*6*	
3		*9*	*12*
	8	***	*16*

* = _______

D3. Use the pictograph.
How much more money was raised in May than March? _____

Money Raised by Bake Sales

March	$$$$
April	$$
May	$$$$$
Key: $ = 5 Dollars	

N4. You plan to use 25 lbs. Of pork and 38 lbs. of beef when you make chili. How many lbs. Of meat will you use?

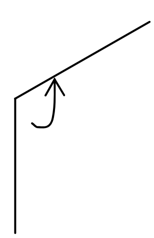
____

A4. Write the number sentence that is missing from the fact family.

9 + 5 = 14
5 + 9 = 14
14 - 5 = 9

____

G4. Choose if the angle is <, >, or = a right angle



a) <
b) >
c) =

M4. Write <, >, or =.

10 in. $\boxed{}$ 2 ft.

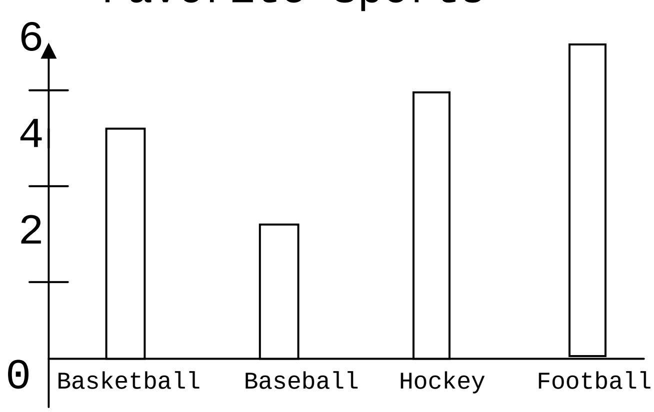
P4. Write the product that the symbol stands for.

1			
	4		*8*
		9	
	8		*

* = ________

D4. Use the bar graph.

Favorite Sports



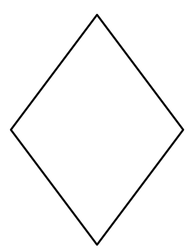
How many more students like football than baseball? _____

N5. Janet wants to buy 4 notebooks for school. Each notebook cost $1.50. How much will she spend? _______

A5. Choose <, >, or =.

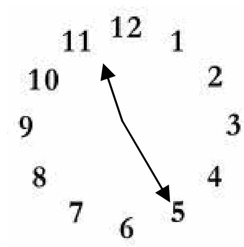
398 $\square$ 276

G5. Find the total number of lines of symmetry for the figure.



a) 2
b) 3
c) 4

M5. Write the time.



____

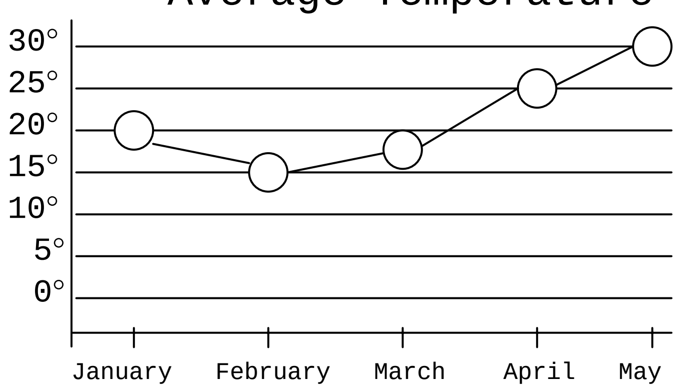
P5. Write the product that the symbol stands for.

1				
	4			*10*
	6			*15*
	8		*	*20*
		15		

* = ________

D5. Use the line graph.
How much colder is it in February than it is in May? ____________

Average Temperature



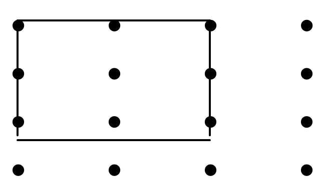
N6. Solve the problem. Decide the remainder.

A football costs $2.50. You have $8.00. How much money will you have left after you buy as many footballs as you can? _______

A6. Choose <, >, or =.

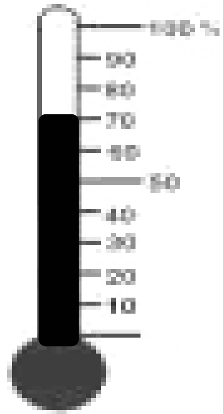
2517 ☐ 3208

G6. Choose how many.



a) 4 square units
b) 5 square units
c) 6 square units

M6. Write the temperature.



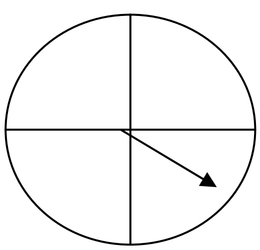
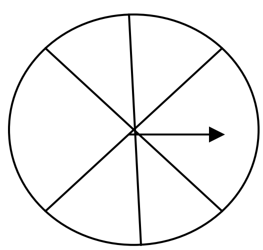
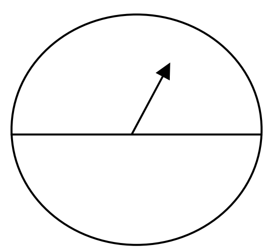
____

P6. Write the product that the symbol stands for.

	2			*5*
		9		*
	10			*25*

* = ________

D6. Use the spinners.

a)  b)  c) 

If you could win a game by stopping on a shaded area most often, which spinner would you want to use?

Fourth Grade
Math Assessment

Teacher____________________________
School_____________________________
Student____________________________
Date_______________________________

N1. Choose the sum.

13 + 66 + 400 + 267

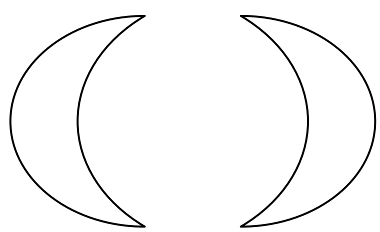
a) 983
b) 647
c) 746

A1. Choose the number to complete the sentence.

(2+3) + 10 = 2 + (+ 10)

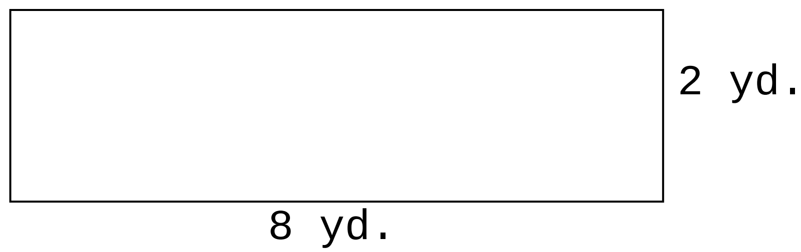
a) 2
b) 3
c) 5

G1. Choose flip, turn or slide for the set of figures.



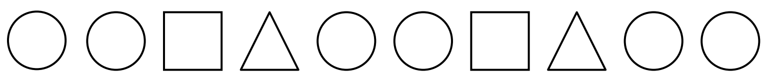
a) flip
b) turn
c) slide

M1. Choose the perimeter.



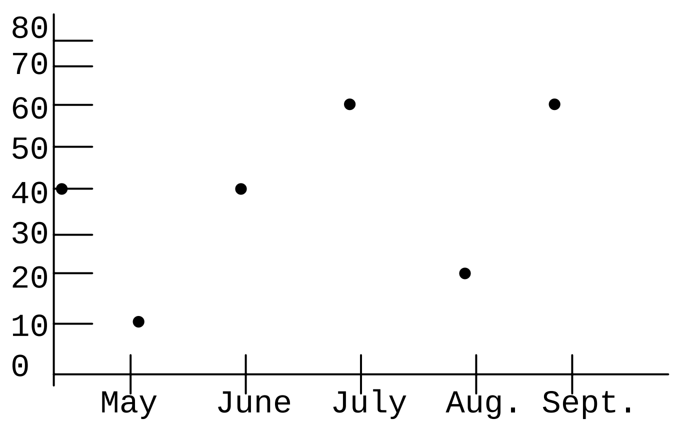
a) 20 yd.
b) 10 yd.
c) 12 yd.

P1. The beads in the necklace follow this pattern.



What shape is the thirteenth bead? _____

D1. Use the graph to find the answer:
How many people rented bicycles in July? ________



N2. 400 – 217

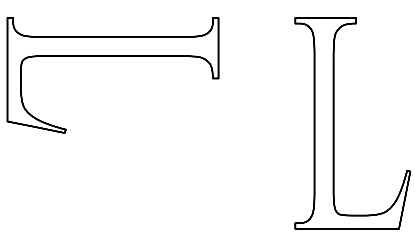
a) 283
b) 183
c) 138

A2. Choose the number to complete the sentence.

(14 +) + 4 = 14 + (15 + 4)

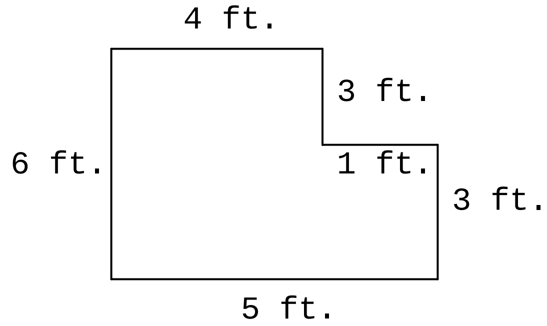
a) 7
b) 14
c) 15

G2. Choose flip, turn or slide for the set of figures.



a) flip
b) turn
c) slide

M2. Choose the perimeter.



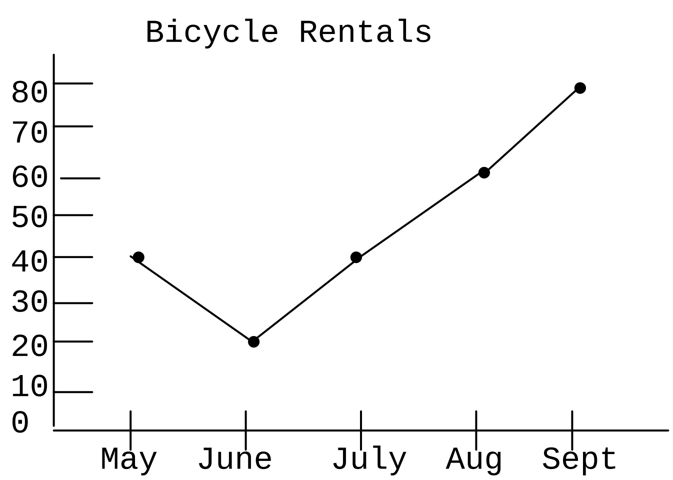
a) 22 ft.
b) 18 ft.
c) 21 ft.

P2. Write the next 2 numbers in this pattern.

4, 3, 7, 6, 13, 12, 25, 24

________ , ________

D2. Use the line graph to find the answer.
How many more people rented bicycles in August than in June? ________



N3. 4.8 – 3.0

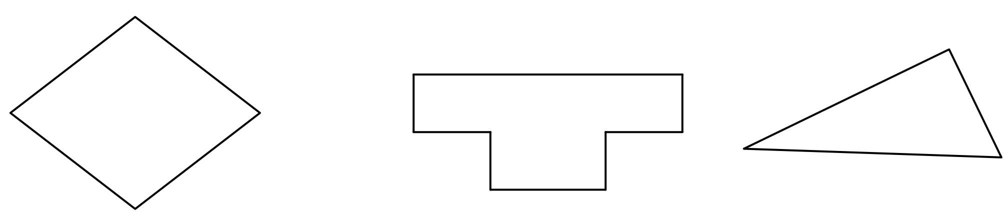
a) 1.8
b) 1.0
c) 2.8

A3. Choose the best estimate.

22 + 63

a) 70
b) 80
c) 90

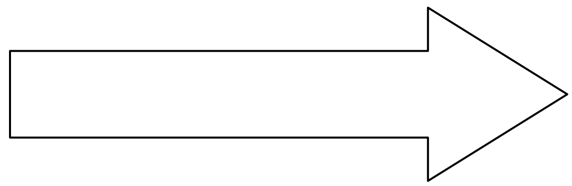
G3. Use the figures to find the answer.



a) b) c)

Which figure has 6 vertices? ________

M3. Measure the length to the nearest inch.

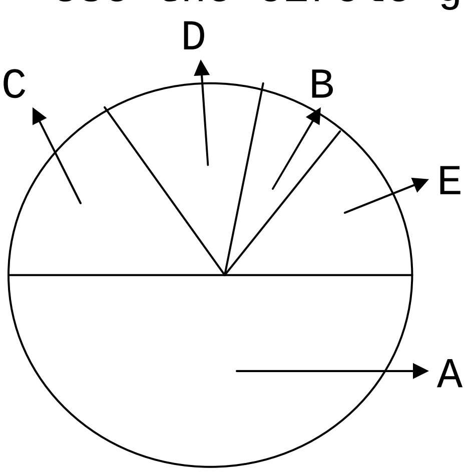


a) 1 1/2 inches
b) 2 1/4
c) 2 inches

P3. Write the next two fractions in this pattern

1/3, 3/6, 5/9, 7/12, ____, ____

D3. Use the circle graph to find the answer.



The greatest number of students like to eat pizza. Which section of the circle would you label as pizza?

a) A
b) B
c) C
d) D
e) E

N4. At summer camp, there are 4 beds along each wall. Each bed is 6 feet long. Between each bed is a 1 foot space. How long must the wall be?

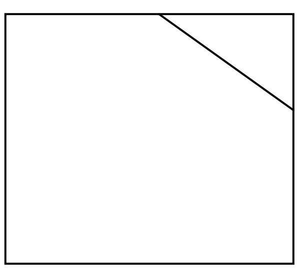
a) 27 feet long
b) 29 feet long
c) 26 feet long

A4. Complete the equivalent fraction. Write the missing number.

$$\frac{2}{3} = \frac{\square}{9}$$

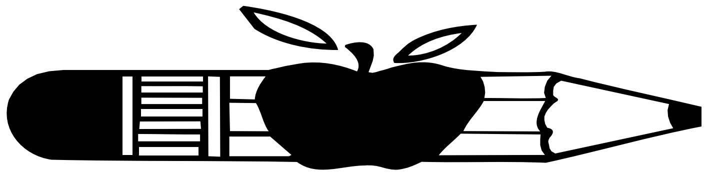
____

G4. Is the line a line of symmetry for the figure? Write yes or no.



____

M4. Measure the length to the nearest centimeter.



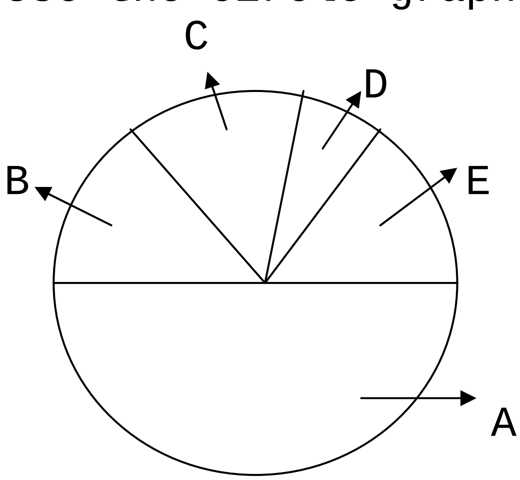
a) 8 centimeters
b) 9 centimeters
c) 10 centimeters

P4. Write the decimals in order from least to greatest.

34, 33.9, 34.1

________, ________, ________

D4. Use the circle graph to find the answer.



The fewest number of students like green beans. Which section of the circle would you label green beans?

a) A
b) B
c) C
d) D
e) E

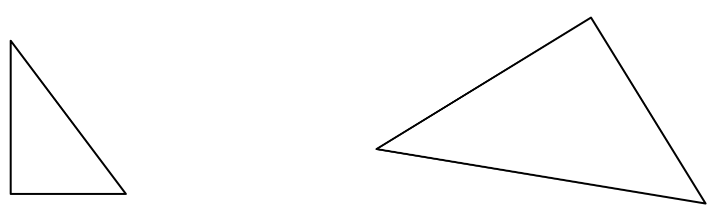
N5. You buy 6 pens for $1.25 each and 3 pencils for $.75 each. You hand the clerk a $10 bill. How much change should you get back?

a) $1.25
b) $.25
c) $.75

A5. Compare. Write <, >, or =.

$$\frac{1}{4} \quad \square \quad \frac{5}{8}$$

G5. Are the figures congruent? Write yes or no.



____

M5. Complete.

6 ft. = $\boxed{}$ yds.

a) 2
b) 3
c) 4

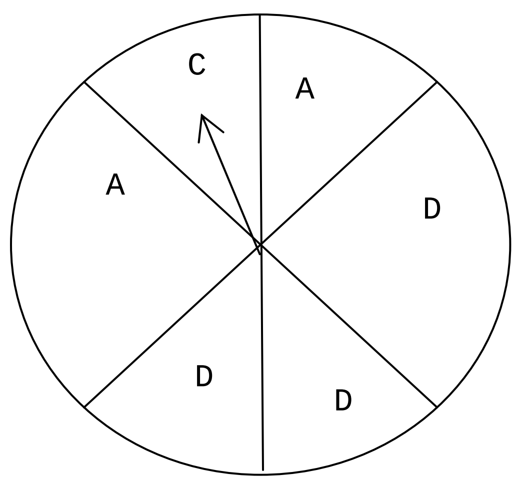
P5. Write the next 2 numbers in the pattern.

2.1, 2.2, 1.9, 2.0, 1.6, 1.7, ____, ____

D5. Use the spinner to choose the answer.

-What are the chances of the spinner stopping on C?

a) 1 out of 6
b) 2 out of 6
c) 3 out of 6

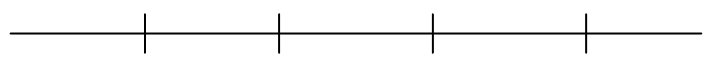


N6. You have 3 cards numbered 2, 3, and 4. How many 2 digit numbers can you make with these cards?

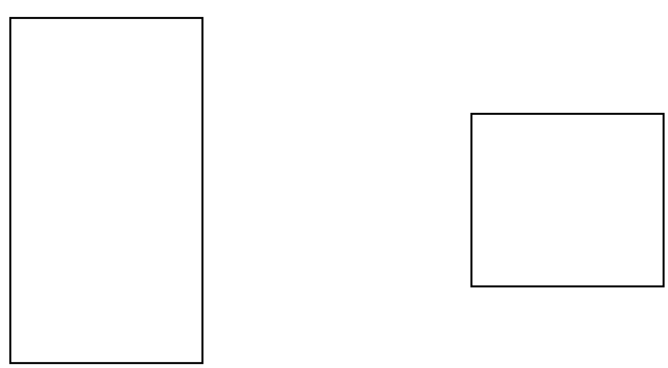
a) 6
b) 9
c) 3

A6. Write the fractions in order from least to greatest.

$$\frac{1}{2},\ \frac{3}{5},\ \frac{1}{10},\ \frac{4}{5}$$



G6. Are the figures similar?
Write yes or no. _______



M6. Complete.

3 kg = $\boxed{}$ g

a) 30
b) 300
c) 3000

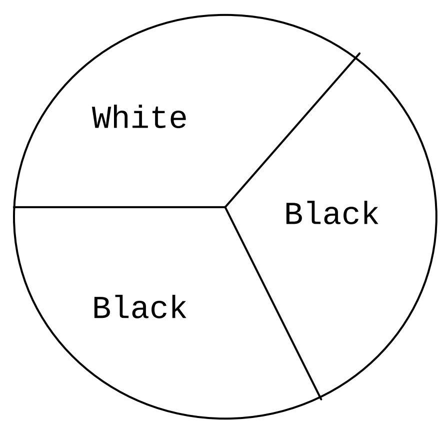
P6. Write the next 2 numbers in the pattern.

8.1, 8.2, 7.9, 8.0, 8.1, 7.8, 7.9,

________, ________

D6. Use the spinner to choose the answer.
-What are the chances of the spinner stopping on white?

a) 1 out of 3
b) 1 out of 4
c) 2 out of 3



Fifth Grade
Math Assessment

Teacher___________________________
School____________________________
Student___________________________
Date______________________________

N1. Choose the least common multiple of 6 and 15.

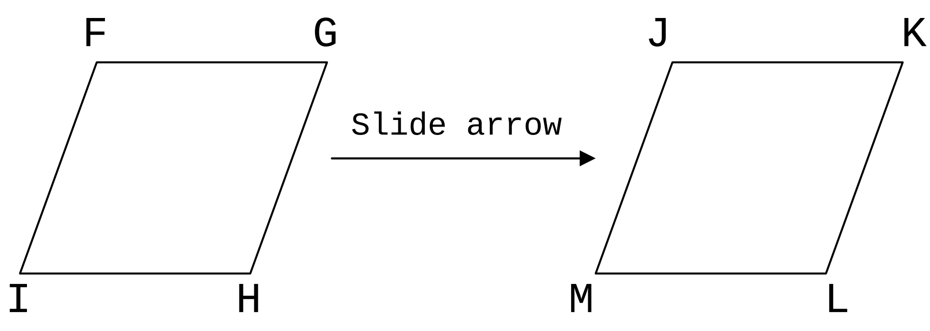
a) 3
b) 9
c) 21
d) 30

A1. Choose the value of each expression.

K + 5 if k = 13

a) 20 b) 18 c) 17

G1. Use the figure.
∠ IFG in figure IFGH is congruent to which angle in the slide image? ______

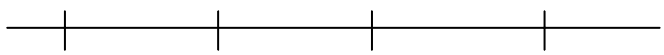


M1. 2/9 + 7/9 =

a) 0
b) 1
c) 9

P1. Order the set from least to greatest.

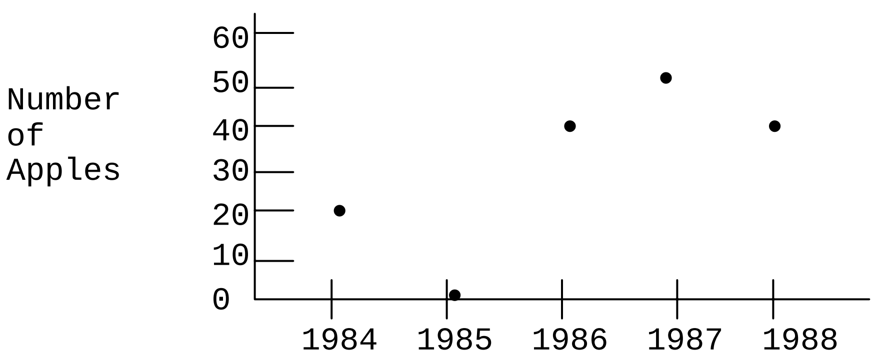
876.2, 826.7, 862.7, 827.6



D1. Use the graph.

How many apples did John pick in 1984, 1985 and 1987 combined?

____



N2. Choose the product.

$3.08 * 8 =

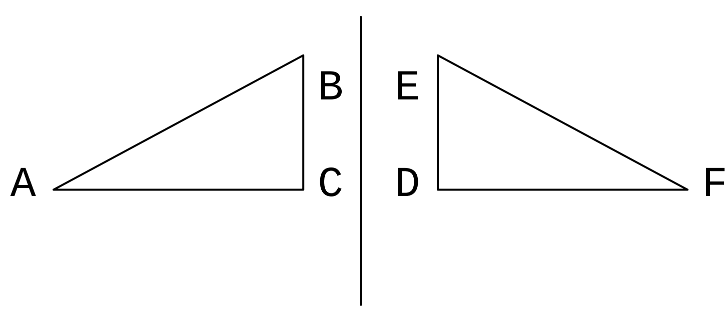
a) 20.64
b) 22.64
c) 24.64
d) 26.64

A2. Choose the value of each expression.

H – 6 =, if H = 13

a) 7 b) 5 c)19

G2. Use the figures.



Flip Line

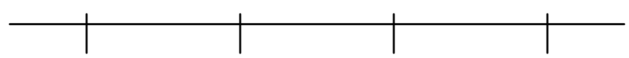
$\overline{AC}$ in figure ABC is congruent to which segment in the flip image? ________

M2. 18/20 - 9/10 =

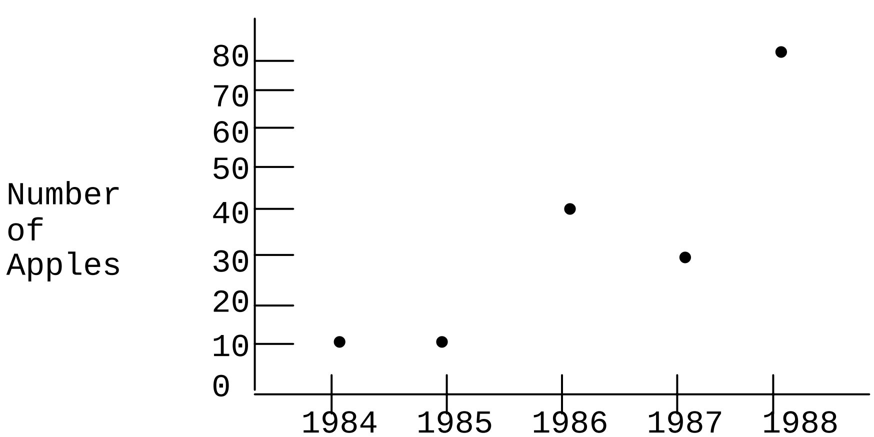
a) 0
b) 9/10
c) 9/20

P2. Order the set from greatest to least

0.73, 1.07, 0.86, 0.66



D2. Use the graph.



Between which two years did the number of apples John picked increase the most?
____

N3. Choose the quotient.

$$9 \overline{)270}$$

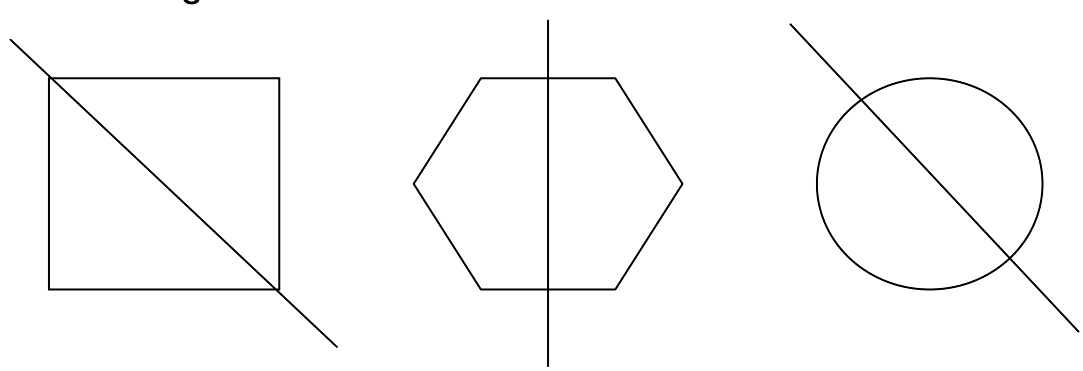
a) 3
b) 33
c) 23
d) 30

A3. The Sportswear catalog company adds a $5.00 shipping charge to the cost of a clothing order.

-If C stands for the cost of an order, what expression tells the final price?

a) C - 5
b) C + 5
c) 5 - C

G3. Use the figures to answer.



Which figures show a line of symmetry?

M3. 11 3/7 - 9 1/2 =

a) 1 13/14

b) 2 1/7

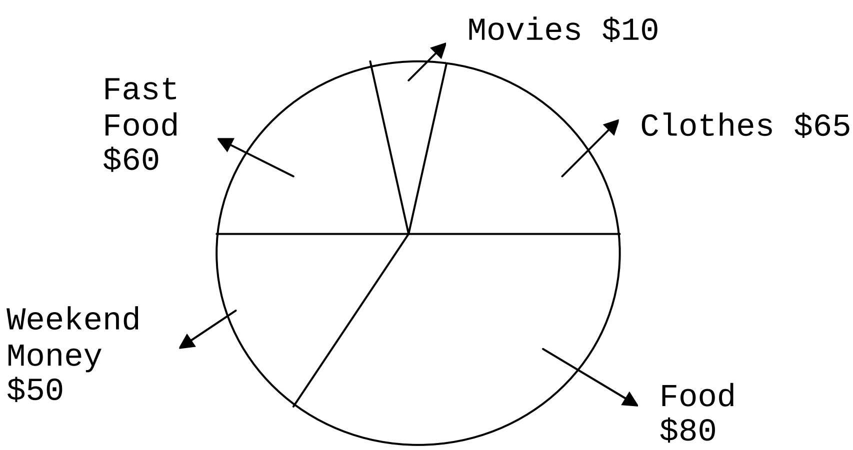
c) 2 13/14

P3. Find the median of the set of temperatures.

59°, 64°, 75°, 78°, 59°

____

D3. Use the pie chart.



Jane's Monthly Expenses

How much money did Jane spend on food altogether?

____

N4. On the average, a taxi driver travels 209 kilometers per day. How far does the taxi driver travel in 5 days?

____

A4. The Sportswear catalog company adds a $5.00 shipping charge to the cost of a clothing order.

-If Joanna orders $22.75 worth of clothes from the catalog, what is her final cost?

a) $24.50
b) $15.75
c) $27.75

G4. Use the figure to answer.

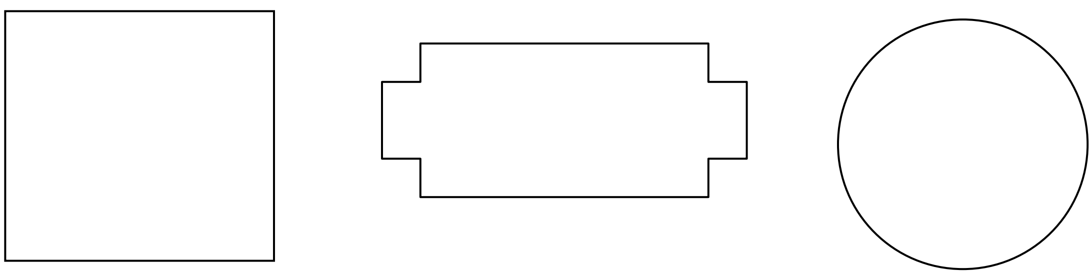


Figure A   Figure B   Figure C

Which two figures have more than one line of symmetry? ________

M4. $6\ 3/5 + 4\ 2/5 =$

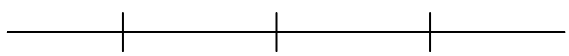
a) $9\frac{1}{10}$

b) $10\frac{1}{10}$

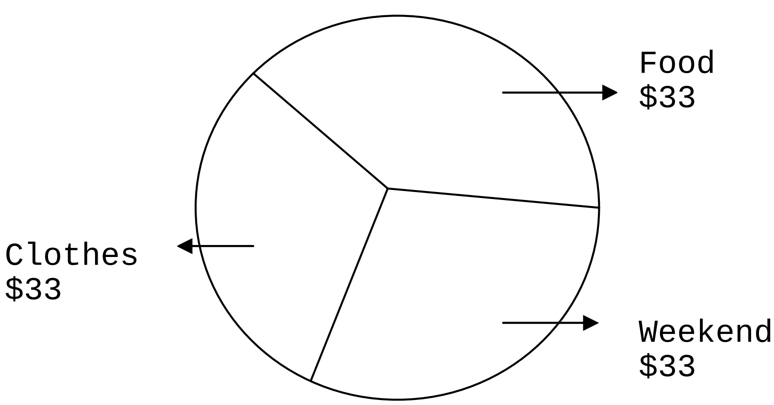
c) $11\frac{1}{10}$

P4. Write the numbers from least to greatest.

$2\frac{2}{7}$, $1\frac{13}{14}$, $2\frac{5}{14}$



D4. Use the pie chart.



Jane’s monthly expenses

What fraction of Jane’s monthly expenses is spent on clothes?

____

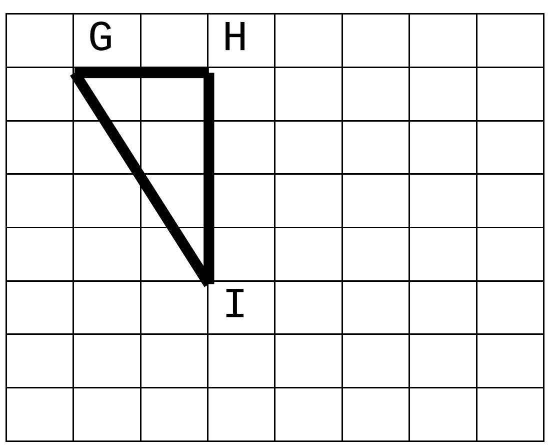
N5. Joe wants to buy one can of soda. The market charges $2.04 for a 6-pack of soda. How much will one can of soda cost Joe?

____

A5. John is reading books for his school. His mother gives him $.10 for each book he reads. If b stands for the number of books John reads, how much money will he get from his father?

a) 5 * b
b) 10 * b-5
c) 10 * b

G5. Use the grid and a ruler to answer.



What type of triangle is GHI?

____

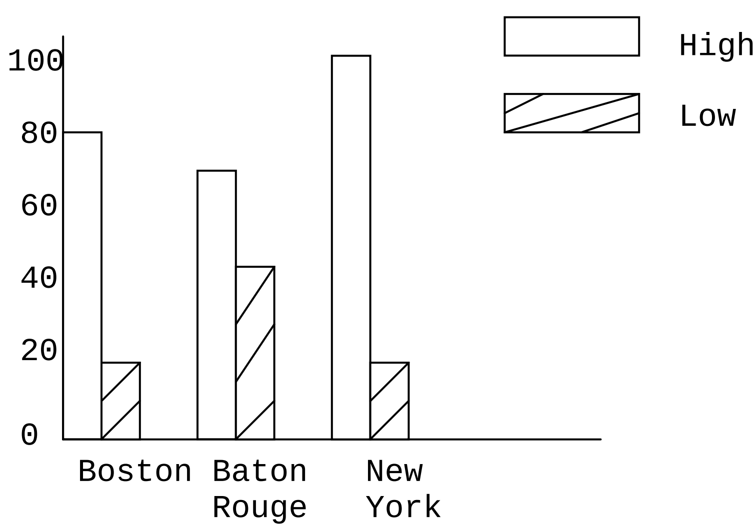
M5. You have 45 inches of rain in 1999, 36 inches of rain in 2000, 42 inches in 2001 and 37 inches of rain in 2002. What is the mean amount of rain?

a) 50
b) 40
c) 160

P5. Write the numbers from greatest to least.

$3\frac{6}{8}$, $3\frac{2}{10}$, $4\frac{1}{12}$

D5.



Which city has the largest range in temperature?

____

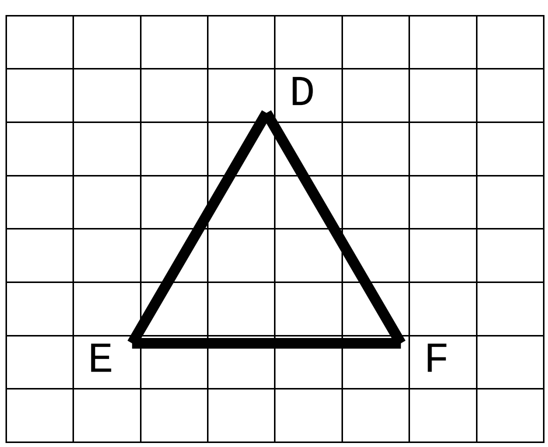
N6. Sandra's bus ride takes 2/3 of an hour. One morning she spent 1/3 of her time reading. How many minutes did Sandra read?

a) 13 minutes
b) 23 minutes
c) 33 minutes

A6. Baseball caps cost $7.00 each. If n stands for the number of caps that Mrs. Brown buys her children, how much money will she spend?

a) 7 * n
b) 4 * n
c) 7 * n-4

G6. Use a grid and a ruler to answer.



What type of triangle is DEF?

____

M6. What is the median of 25, 26, 27, 28, 29?

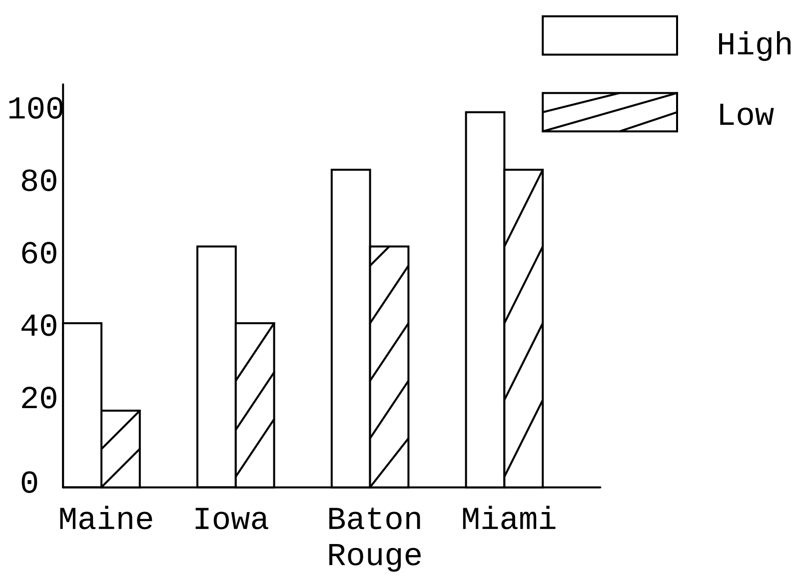
____

P6. Decide if the ratio is equivalent.

6 to 4 and 9 to 6

a) yes
b) no

D6. Use the graph.

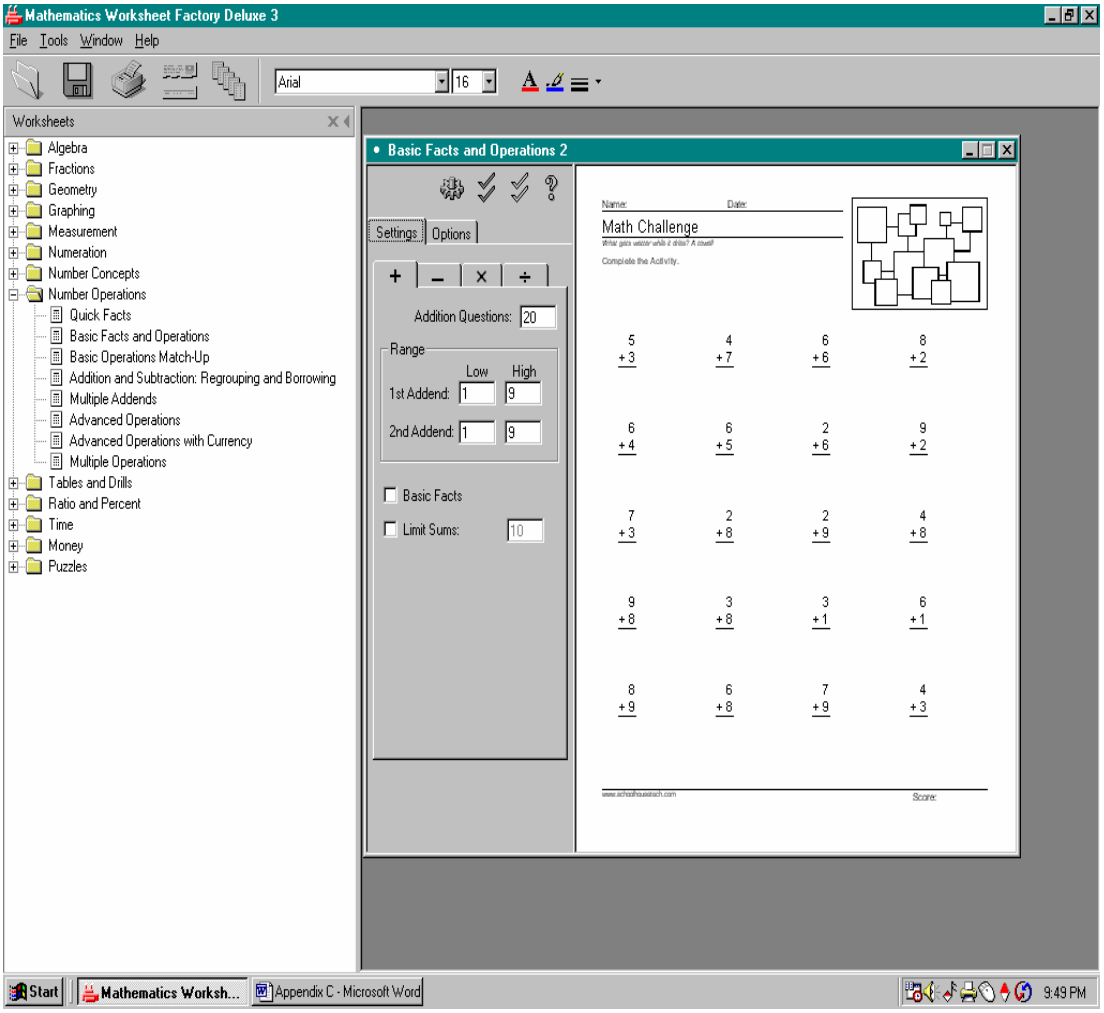


What was the mean temperature for Miami?

____

Appendix C

Appendix C. Basic Skill Worksheet, Sums to 18



Appendix D

Appendix D. Teacher Rating of Student Performance and Expected Final Grade.

Name & Grade	Computation						Application					
	Mastery	Instr .	Satisf	Some diff	Frustra	Cannot perform	Mastery	Instr .	Satisf	Some diff	Frustra	Cannot perform
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												
Student Grade												

*Five categories: Mastery, Instructional, Satisfactory, Some difficulty, Frustrational, Cannot perform

Teacher Name____________________ School ____________________
Date ____________________ Grade ____________________

Appendix E

Appendix E. Assessment Rating Profile.

Assessment Rating Profile – (ARP-15)

Directions: Please rate the math assessment along the following dimensions. Please circle the number which best describes your agreement or disagreement with each statement.

	Strongly Disagree	Disagree	*Disagree Slightly*	*Slightly Agree*	*Agree*	*Strongly Agree*
1. This would be an acceptable assessment for assessing overall student performance in math.	1	2	3	4	5	6
2. Most teachers would find this assessment appropriate for assessing overall math performance.	1	2	3	4	5	6
3. This assessment should prove effective in monitoring student performance across the year.	1	2	3	4	5	6
4. I would suggest this assessment to other teachers who want to assess math performance.	1	2	3	4	5	6
5. The student's math skills can be monitored using this assessment.	1	2	3	4	5	6
6. Most teachers would find this assessment suitable for assessing math performance.	1	2	3	4	5	6
7. I would be willing to use this assessment in the classroom setting.	1	2	3	4	5	6
8. This assessment would *not* result in negative side effects for the child.	1	2	3	4	5	6
9. This assessment would be appropriate for a variety of children.	1	2	3	4	5	6
10. This assessment is consistent with those I have used in classroom settings.	1	2	3	4	5	6
11. The assessment was a fair way to assess the content standards	1	2	3	4	5	6
12. This assessment is reasonable for the content standards described.	1	2	3	4	5	6
13. I liked the items used in this assessment.	1	2	3	4	5	6
14. This assessment is a good way to assess math performance.	1	2	3	4	5	6
15. Overall, this assessment would be beneficial for a child.	1	2	3	4	5	6

Appendix F

Appendix F. Directions and Scripted Instructions for Single Skill and Multiple Skill Math and Maze Assessments.

Procedures for Math Assessments

1. Worksheets are passed out face down on student's desks and students are informed they are not to turn over the worksheets until they are told to do so.
2. Instructions are given: **"Please write your first and last name on the back of the worksheet, and your teacher' s name."** Ask students to look at you when they are finished and allow time to write.
3. Say, **"This is a math worksheet. When I say 'go', turn your worksheets over and begin solving the problems. Start form the left side of the page on the top row and work your way across the page to the right without skipping any problems. If you do not know the answer to a problem, take your best guess and go to the next one. You will not be penalized for putting down the wrong answer, so do not skip any questions. Are there any questions?**" Set timer for two or five minutes, depending on the probe.
4. Say, **"Start".** Start timer. Wait two or five minutes.
5. When the timer goes off, say, **"Stop now, put your pencil down."**
6. Mark where the students stopped.
7. Say, **"You may now complete the worksheet. Put an answer for every problem. If you don not know the answer, take your best guess." "Put your paper I the air when you are finished."**
8. Collect finished worksheets.

Procedures for Maze Assessments

1. Worksheets are passed out face down on student's desks and students are informed they are not to turn over the worksheets until they are told to do so.
2. Instructions are given: **"Please write your first and last name on the back of the worksheet, and your teacher' s name."** Ask students to look at you when they are finished and allow time to write.
3. Say, "**This is a reading exercise. I want you to read the passage up to blank line. Located next to the blank you will find three words. Circle the word that completes the sentence. Continue circling the words that complete the sentences until the timer goes off. Do you have any questions?**
4. Say, **"Start".** Start timer. Wait two minutes.
5. When the timer goes off, say, **"Stop now, put your pencil down."**
6. Mark where the students stopped.
7. **"You may now complete the worksheet. Put an answer for every problem. If you don not know the answer, take your best guess." "Put your paper I the air when you are finished."**
8. Collect finished worksheets.

Appendix G

Appendix G. Procedural Integrity Check List.

Place a check (√) indicating complete, Not Applicable (NA) or Not None (ND) next to each step.

Math:

1. Students were instructed to write their name and their teacher's name. _____
2. Students were read directions._____
3. Students were given two (or five) minutes to complete the fluency assessment measure._____ or allowed to complete applications assessment___
4. Probes were collected at the end of the two minutes or when the students finished._____

Maze:

1. Students were instructed to write their name and their teacher's name. _____
2. Students were read directions. _____
3. Students were given two minutes to complete the Maze passages._____
4. Passages were collected at the end of the two minutes._____

Teacher Name	________________________
Grade	________________________
School	________________________
Date	________________________
Instructor (LSU)	________________________
Rater (LSU)	________________________

Vita

James Edward Connell is a candidate for the degree of Doctor of Philosophy in the school psychology program at Louisiana State University. He graduated from Temple University with a Bachelor of Arts degree in psychology and did post baccalaureate work at Children's Seashore house and the University of Pennsylvania under the supervision of Dr. F. Charles Mace. He also taught second grade as a full-time substitute teacher at Polk Elementary School in East Baton Rouge Parish under the supervision of Principal Lee Dixon. He earned a masters' degree under the supervision of Dr. Joseph C. Witt and is currently following a doctoral program under the supervision of Dr. Witt. at Louisiana State University. He expects to graduate with the Doctor of Philosophy degree in August, 2005.